UNIVERSITY OF NAIROBI

SCHOOL OF COMPUTING AND INFORMATICS

Model Ensembles for Predictive Drought Severity and Drought Effects Monitoring using Remote Sensing & Socio-Economic data

Chrisgone Adede Otieno

P80/98347/2015

Supervisors:

Prof. Robert Oboko

Prof. Peter Waiganjo Wagacha

JUNE, 2020

A thesis submitted in fulfilment of the requirements for the award of the degree of Doctor of Philosophy in Computer Science

Declaration

This thesis is my original work and has not been presented for a degree in any other university for examination.

Signature: ______________________________Date: ____________________

Chrisgone Adede Otieno

This thesis has been submitted for examination with our approval as university supervisors.

Signature: ______________________________Date: ____________________

Prof. Robert Oboko

School of Computing and Informatics, University of Nairobi

Signature: ______________________________Date: ____________________

Prof. Peter Waiganjo Wagacha

School of Computing and Informatics, University of Nairobi

Dedication

To the scientists that realize how daring it is to create models and deploy them in operational environments in which resources follow decisions made based on the model outputs.

Acknowledgement

The PhD journey remains one of the most exciting not only for the student but also for the entire team involved. Even though not a straight one, the journey is one that pans out well with supreme guidance and protection of God.

This work made it to the finish line due to the supervisory assistance of both Prof. Robert Oboko and Prof. Peter Wagacha in shaping the research and thought processes. It is on many occasions that Prof. Oboko and I scheduled meetings for 0600HRS in students' halls to review the progress of this work! Prof. Oboko and Prof. Wagacha walked with me through the PhD process providing required critical review of the work. It was beautiful learning from professionals whose mentorship was nothing but the best!

My gratitude also goes to Prof. Clement Atzberger of the University of Natural Resources and Life Sciences, Vienna (BOKU). The mentorship from Prof. Clement in remote sensing through the collaboration with the National Drought Management Authority (NDMA) has been an eye-opener. I appreciate the study period I spent in the BOKU Labs learning R for remote sensing. Prof. Clement's tutelage on writing journal papers and the creation of ordered concepts towards publishing in top of the field journals was invaluable. The editorial work Prof. Clement undertook on our publications and the critical review of the drafts together with his guidance on how to respond to editorial questions even when they seemed daunting are indescribably appreciated.

To my friend and work colleague Julius Mburu, countless times, you made follow-ups on this work. You are that voice in my head that gave me justifications to proceed. I recall the discussion with Luigi Luminari when he gave his opinion on problems in drought monitoring at the NDMA that required research. That discussion led to this fulfilling journey.

I will not forget to thank my family members who have been supportive even on many occasions when my absence was a necessity. My wife, Jane Arwa for her support in many ways- both in logistics and moral. Daughter Amani who asked why school work has to result in a lot of write-up: hopefully, you will come to understand. To Mitchelle, Gabrielle and Ella, your collective patience over the study period is appreciated. To my siblings: Rockliffe, Kenneth, Brian and Nova, thanks for walking with me through this journey. Finally, to my Dad for always thinking this was worth the investment.

TABLE OF CONTENTS

List of Tables

List of Figures

List of Acronyms

AC	-	Agreement Correlation
ANN	-	Artificial Neural Networks
ASALs	-	Arid and Semi-Arid Lands
AVHRR	-	Advanced Very High-Resolution Radiometer
BOKU	-	University of Natural Resources and Life Sciences, Vienna
CHIRPS	-	Climate Hazards Group InfraRed Precipitation with Station data
CRISP-DM	-	Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology
DCF-P	-	Drought Contingency Fund Project
DCM	-	Drought Cycle Management
DEWS	-	Drought Early Warning System
DRM	-	Drought Risk Management
DV	-	Data Visualization
ECDF	-	Empirical Cumulative Distribution Function
EDE	-	Ending Drought Emergencies
EWS	-	Early Warning System
fAPAR	-	Fraction of absorbed photosynthetically active radiation
FEWSNET	-	Famine Early Warning Systems Network
GEOSS	-	Global Earth Observation System of Systems
GoK	-	Government of Kenya
HSNP	-	Hunger Safety Net Programme
ICPAC-	-	IGAD Climate and Prediction Center
IGAD	-	Intergovernmental Authority on Development
ILRI	-	International Livestock Research Institute
KDA	-	Knowledge Discovery Approach
KDD	-	Knowledge Discovery from Databases
LANDSAT	-	Land Satellite
ML	-	Machine Learning
MODIS	-	Moderate Resolution Imaging Spectroradiometer

MUAC	-	Mid-Upper Arm Circumference
NASA	-	The National Aeronautics and Space Administration
NDMA	-	National Drought Management Authority
NDVI	-	Normalized Difference Vegetation Index
NOAA	-	National Oceanic and Atmospheric Administration
PIT	-	Point in Time
PPT	-	Precipitation
PRD	-	Prediction Research Design
RCI	-	Rainfall Condition Index
RFE	-	Rainfall Estimates
RS	-	Remote Sensing
RSD	-	Remote Sensing Data
SE	-	Socio-Economic
SED	-	Socio-Economic Data
SEMMA	-	Sample, Explore, Modify, Model, and Assess (SEMMA) methodology
SMTP	-	Second Medium-Term Plan
SPOD	-	Sum of Potential Difference
TAMSAT	-	Tropical Applications of Meteorology using SATellite
UNISDR	-	The United Nations Office for Disaster Risk Reduction
VCI	-	Vegetation Condition Index
VS2030	-	Vision 2030
ZVI	-	Standardized Vegetation Index

Operational Definition of Terms

For a better understanding of this thesis, the key terms in the study are given the following operational definitions:

Drought Severity: Denotes the vegetation deficit as based on a reference threshold and typically have values between 0 and 100 in this study with a reference sent at 35 to denote non-drought conditions in the classification of drought. Despite the implied similarity between drought severity and drought intensity which is the ratio of vegetation deficit and the duration of the drought given that we a monthly frequency of drought monitoring, we adopt drought severity to imply both in the context of this study.

Predictive drought monitoring: Is used to imply an operational drought monitoring system that has both aspects of drought monitoring and forecasting of future drought conditions. For every given frequency of monitoring, actual drought conditions are provided as well as a forecast of future conditions.

List of Equations

List of Publications

The study has realized two publications covering the prediction of drought severity. Each of the publications outlines the role of the authorships and therefore illustrate their appropriateness for identification as outputs of this study. The publications are:

- Adede, C., Oboko, R., Wagacha, P. W., & Atzberger, C. (2019). A Mixed Model Approach to Vegetation Condition Prediction Using Artificial Neural Networks (ANN): Case of Kenya's Operational Drought Monitoring. *Remote Sensing*, *11*(9), 1099. https://doi.org/10.3390/rs11091099

  The *Remote Sensing* journal had an impact factor of 4.118. The paper was submitted on the 23rd of January, 2019 and Accepted on 2nd May, 2019. The paper was subjected to 4 reviewers over 2 rounds of reviews. The paper built multiple drought severity prediction models and validated all the assumptions that were made in the main study. The main achievement was the success in the validation of all the assumption for the main study especially those that realized a reduced model space to overcome the possible problem of combinatorics.

- Adede, C., Oboko, R., Wagacha, P. W., & Atzberger, C. (2019). Model ensembles of artificial neural networks and support vector regression for improved accuracy in the prediction of vegetation conditions and droughts in four northern Kenya counties. *ISPRS International Journal of Geo-Information.* https://doi.org/10.3390/ijgi8120562

  Submitted to *ISPRS International Journal of Geo-Information* on the18th of September 2019, the paper was accepted on the 28th of November, 2019 after two rounds of reviews from two reviewers and one round from a third reviewer. The journal had an impact factor of 1.840 at the time of publication. The paper explores model ensembling as an approach to realize highly predictive models. It investigated the performance of three model ensembling approaches as applied on multiple ANN and SVR models.

Abstract

Droughts, with their increasing frequency of occurrence, especially in the Greater Horn of Africa (GHA), continue to negatively affect lives and livelihoods. For example, the 2011 drought in East Africa caused massive losses documented to have cost the Kenyan economy over US$ 12.1 billion. Consequently, the demand is ever-increasing for ex-ante drought early warning systems with not only the ability to offer drought forecasts with sufficient lead times but that are both stable and are of high bias. In this study, we build predictive models one month ahead for both drought severity and drought effects. Vegetation condition index aggregated over 3 months (VCI3M) and nutrition of children below 5 years as indicated by Mid-Upper arm circumference (MUAC) are used as the proxy variables for drought severity and drought effects respectively. We present the performance of both homogeneous and heterogeneous model ensembles in the prediction of drought severity and drought effects using the study case techniques of artificial neural networks (ANN) and support vector regression (SVR). For each of the homogeneous and heterogeneous model ensembles, we investigate the performance of three model ensembling approaches of simple averaging, ranked weighted averaging and model stacking. Applying the approach of over-produce then select, the study used 17 years of remote sensing data and 10 years of socio-economic data to build 244 individual ANN and SVR models from which 111 models were selected for the building of the model ensembles. The results indicate the superiority of the heterogeneous model ensembles to both homogeneous model ensembles and individual champion models. Model stacking as applied in heterogeneous model ensembles is shown to be superior to both simple average and weighted average ensembles. The heterogeneous stacked model ensemble recorded an R^2 of 0.94 in the prediction of drought severity as compared to an R^2 of 0.83 and R^2 of 0.78 for both ANN and SVR champion models respectively. The superiority of the heterogeneous stacked ensemble is extended to classification in which accuracy of 80% is recorded as compared to 69% and 71% for the ANN and SVR champion models respectively. Additionally, the poor performance of champion models in outlier classes is mitigated on by the use of stacked heterogeneous model ensembles. We conclude that despite the computational resource intensiveness of the model ensembling approach to drought prediction, the returns in terms of model performance is worth the investment, especially given the recent exponential increase in computational power. We nevertheless advise evaluating the use of more techniques in the model ensembles and the building of many more ensembles using divergent ensemble sizes to settle the question of performance of model ensembles fully. To further increase the utility of drought prediction, we also recommend the study of more extended forecasting periods (up to 6 months) and to estimate how much this would degrade the prediction skill of the ensemble models.

Keywords*:* general additive model; drought risk management; early warning system; ensemble; over-fitting; model space reduction; support vector regression.

Chapter 1: INTRODUCTION

1.1 Background

Droughts have been documented to have effects that can be viewed from different perspectives. The social, environmental and economic effects of drought are noted to shape the response of both research and policy. In this section, we introduce drought from the different perspectives including economic, social, environmental, policy and research views of the effects of droughts.

1.1.1 The Economic View

Drought is described by Below, Grover-Kopec & Dilley (2007) and Olang et al. (2013) as one of the greatest impediments to development in Africa due to reliability on rain-fed agriculture and high vulnerabilities as a result of poverty. Much of the continent is dependent on rain-fed agriculture, which makes it particularly susceptible to climate variability (Di Falco & Veronesi, 2013).

Drought economic losses range from those resulting from poor agricultural production of both crop and livestock, loss of revenue from agricultural taxes, poor power production from hydro-power dams, interference with transportation waterways, timber and lumbering losses and strain on institutions that offer both credit and credit risk insurance (Ding, Hayes & Widhalm, 2011; Wilhite & Glantz, 1985).

Several economic impacts of drought are documented in literature. The World Bank (2011) documents that the prolonged Kenyan drought of 2008-2011 resulted in combined damages and losses of up to US$ 12.1 billion by 2011. The Kenya drought is documented to have left 3.7 million people faced with hunger. Howitt, Medellin-Azuara & Lund (2014) documents that the overall effect of 2014 California Central Valley drought is estimated at nearly US$ 2.2 billion with about 17,100 full time and seasonal job losses. Adams et al. (2002) suggest that the 2002-2003 Australia drought and its flow-on effects would have had up to a 1.6% loss in gross domestic product (GDP). In the Kenya, Australia and California droughts, for example, multiple sectors were affected including but not limited to crop agriculture, livestock agriculture, water, nutrition, education and security.

1.1.2 The Social View

Socially, drought has several effects ranging from psychological impacts as a result of the loss of key assets, especially among pastoral communities. Other social effects include: health and nutrition problems resulting from limited access to water and food, increased threats of fire and need for constant migrations that lead to interference with family setups. Wilhite, Svoboda & Hayes (2007), for example, documents the complexity of drought impacts and identifies conflicts especially in the access to water resources. At the very extreme, droughts have resulted in deaths of people and animals in addition to the increase in workloads on the society, food insecurity and the possible impacts of malnutrition (Keshavarz, Karami & Vanclay, 2013).

In the East African context, and in recent times, populations and communities are increasingly faced with the probability of disasters arising from drought as a hazard. The droughts tend to be more frequent, longer and more severe in East Africa and the Greater Horn of Africa (GHA) as documented in Gebremeskel et al. (2019). These droughts have had the result that communities have their livelihoods disrupted to the extent that they are then unable to use own resources to cope with the loss consequences.

Popularly, the social and economic impacts of drought are always lumped together with the use of the terminology "socio-economic" impacts as both hold the view of the effects of droughts on lives and livelihoods. The use of the socio-economic terminology is for example in Musolino, de Carli & Massarutto (2017) and Chand & Biradar (2017).

1.1.3 The Environmental View

The elements of the environment, including plants, animals, climate, soils, rocks and many others are vastly affected by drought conditions. Droughts, therefore, affect both different aspects of the ecosystem and the environment. In the most, and without guarantee, some these elements recover after droughts. Permanent destruction occurs when desertification and loss of wildlife occurs. The loss of species is for example documented in Kala & Silori (2013) to most common amongst species that have low population sizes coupled with a narrow range of distribution. The destruction of both aquaculture and wildlife habitats is, therefore, a common consequence of drought.

Wind erosion of bare soils, wildlife migration, loss of wetlands, stress to water sources and even depletion of water resources are some of the few impacts of drought on the environment as reviewed in Kala (2017).

1.1.4 The Policy View

There has been an increase in global concern over the ineffectiveness of current approaches to drought risk management (DRM) that have been largely characterized by the crisis management approach. Such approaches are reactive to the occurrence of droughts (Wilhite, Sivakumar, & Pulwarty, 2014). The concern is the need for a structured and proactive approach to DRM. The need for DRM has for example seen initiatives such as the multi-stakeholder High-level Meeting on National Drought Policy (HMNDP) in 2013. The HMNDP aimed to identify science-based actions capable of addressing issues in DRM and to outline possible strategies for enhanced coping capacities (Sivakumar et al., 2014). Strategic frameworks like the Drought Resilient and Prepared Africa (DRAPA) are a direct response to the HMNDP initiative to build an effective DRM and to build drought resilience at multiple levels: continental, regional, national, or local/community levels for Africa.

Directly from the above realization of the need for structured and institutionalized DRM as a result of losses associated with droughts, most countries have adopted policy considerations meant to contain drought impacts. The need for counties to adopt national level drought policies is for example championed in Sivakumar et al. (2011). The key outcome of the policy response to DRM is the reduction of drought disaster risks through structured programmatic activities by countries, regional bodies and even international bodies.

Locally, the Kenya case that was previously characterized by crisis response and uncoordinated management approaches to droughts and their impacts has since seen a policy shift in the three main fronts including; the mainstreaming of drought and drought risk management into development planning, prioritization of drought monitoring and the formulation of relevant policies and institutions for drought risk management. Such initiatives include:

- The 2010 Constitution of Kenya, CoK that outlines in article 43 the citizen's right to be free from hunger and to have adequate food of acceptable quality (Constitution, 2010)
- The establishment of the National Drought Management Authority (NDMA) in 2011 as a specialized institution for DRM with the mandate to coordinate all matters relating to DRM in Kenya (Klisch, Atzberger & Luminari, 2015).
- The inclusion of the Ending Drought Emergencies (EDE) as part of the Second Medium Term Plan (MTP II) of the Kenya Vision 2030 that anchors DRM and the goal of EDE long-term development blueprint for Kenya.

1.1.5 The Research and Technical View

Research in drought has been an ongoing initiative that is gaining popularity with increasing focus and investments. The Drought Cycle Management model (DCM), the classic variant of which is given in Figure 1.1, is the popular drought management model of choice (Oxfam, 2009; Pantuliano & Wekesa, 2008). The DCM views drought as a natural disaster that is both slow in onset and that has effects in phases. Through the DCM (Figure 1.1), drought is characterized to have four phases: Normal, Alert/Alarm, Emergency and Recovery. The current practice in drought monitoring has however seen the extension from the initial four phases to the current five phases as a result of the separation of the drought alert from the alarm stages.

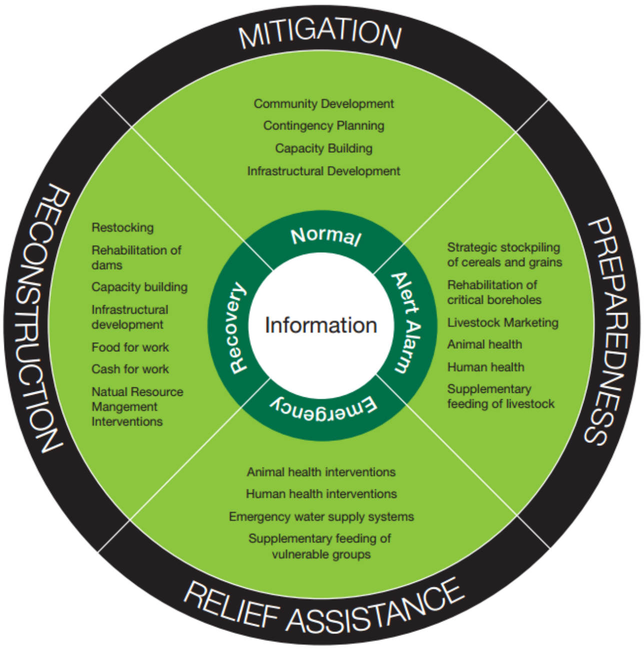


Figure 1.1: Drought Management Cycle (Oxfam, 2009).

Different monitoring systems have different definitions of the drought monitoring phases. An example in case is the Kenyan cases where the National Drought Management Authority (NDMA) that subscribes to the five phases definition outlined above. Even though Figure 1.1 seems to suggest that the transition between the stages is linear, in practical application, the transitions are back and forth between the stages.

The research areas that have emerged out of the DCM model and that are used in the management of drought currently are Drought Early Warning Systems (DEWS), drought preparedness and drought response. These research areas are as briefly discussed:

1. **Drought Early Warning Systems (DEWS)**

   Drought Early Warning Systems (DEWS) are the backbone of drought monitoring and management. The increasing popularity of predictive DEWS is based on their ability to aid stakeholders to react before a crisis occurs especially in the light of increased damages from droughts (Adede et al., 2019b). The implementation and deployment of DEWS is made possible by different information management approaches. Despite the difference in approaches to their implementation, effective DEWS should be punctuated by the ability to assess, communicate and trigger action. Despite their reliance on technology, DEWS should remain accurate, simple, reliable, flexible and timely in the provision of actionable information as documented in Magno et al. (2018) and in Motha, Wilhite & Wood (2011). Information management for drought monitoring should be a continuous undertaking across all the phases of drought.

2. **Drought Preparedness**

Drought preparedness is a wider concept within drought risk management (DRM) that includes drought monitoring and forecasting, vulnerability mitigation, resilience building, impact assessments and response planning (Gutiérrez et al., 2014). Drought preparedness, therefore, involves long-term undertakings of development activities and emergency planning that are aimed at reducing vulnerabilities of communities to drought effects. Drought preparedness is, in effect, the sum-total of pre-disaster, as well as during and after disaster initiatives. Drought preparedness is thus the focal point of the disaster risk reduction framework in IGAD (2007) that is presented in Figure 1.2. Closely related to the concept of drought preparedness is drought risk reduction (DRR) that generally advocates for sufficiency of interventions that enhance local capacities for disaster prevention and emergency preparedness to avoid disasters. The approach of disaster risk reduction is for example adopted in Government of Kenya (2014) for the Kenya Vision 2030's Ending Drought Emergencies Common Programme Framework (EDE: CPF).

3. **Drought Response**

In drought risk management (DRM), drought response is also referred to as drought impact mitigation and is the collection of all efforts that aim to mitigate the impacts of on-going droughts on both lives and livelihoods. A key objective of drought response is the provision of relief to the affected population, especially water, food and health care (FAO, 2019). The limitation of drought response is that the mitigation measures are deployed after drought effects are felt on lives and livelihoods. There is an increasing tendency to minimize investments in drought response with the current paradigm shift from crisis to risk management (Wilhite, 2014).

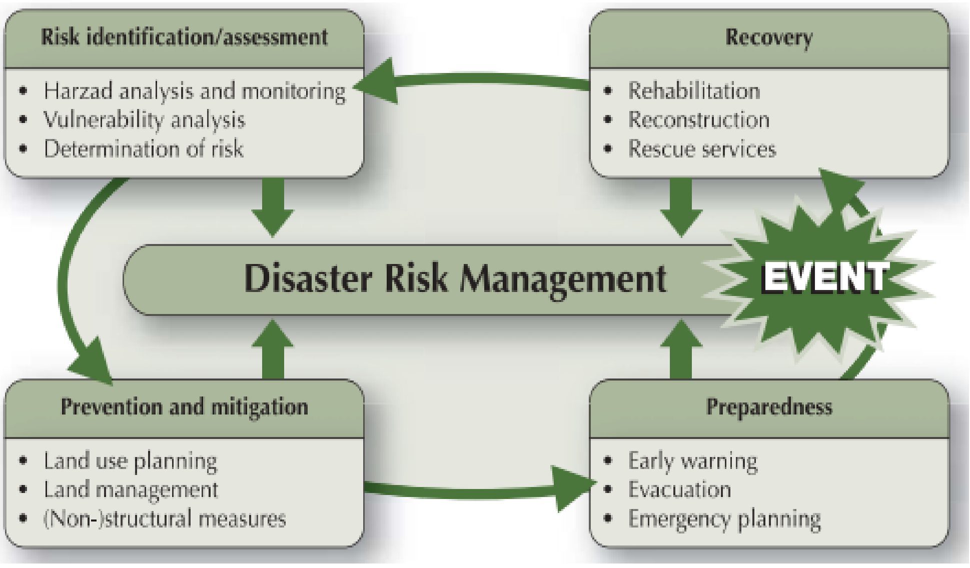


Figure 1.2: Disaster risk reduction framework (IGAD, 2007)

1.2 The Current State of the Art and Future Expectations of Drought Early Warning Systems

The current state of the art is that there is an extensive focus in drought early warning systems for drought monitoring in most countries. Increasingly, permanent institutions for drought monitoring are being established within countries, regionally and even globally. Such institutions for drought risk management include the National Drought Management Authority (NDMA) at the local level in Kenya (Oduor, Swift & Birch, 2014), the United Nations Office for Disaster Risk Reduction (UNISDR, 2012) at the global level and the Intergovernmental Authority on Development (IGAD) (ICPAC, 2007) at the regional level. These local, regional and international institutions are both engaged in and are offering support in the area of drought monitoring.

The current trends in early warning either involves the use of a single indicator/index, mostly sourced from one data source like the univariate cases in Klisch & Atzberger (2016), AghaKouchak & Nakhjiri (2012) and Brown et al. (2015). The use of multiple indicators is only in a few studies and is not as common. Studies that incorporate multiple indices include those in Tadesse et al. (2010), Tadesse et al. (2014) and in Wardlow et al. (2012) that incorporated 11 variables derived from oceanic, environment, climate and satellite data. Incorporation of remote sensing and ground-based data collection approaches have been suggested for use in early warning systems including from the study in Enenkel et al. (2015). Practical implementations or even the possible deployment of such an integrated approach that uses both remote sensing data and ground-truthing data is widely missing from literature.

The current state of the art is domiciled in increased losses from natural disasters in general (Da Silva, 2012). Increase in drought occurrences and damages are also documented (Howitt et al., 2015; UNISDR, 2012; World Bank, 2011). Howitt et al. (2015) disaggregate the losses to include a quantification of the economic impact and job losses of the 2015 California drought. World Bank (2011) documents and quantifies the losses from the 2008-2011 Kenyan droughts at US$ 12.1 billion. There have been, however, counter-arguments to the observed increases in loses from natural disasters. McMullan et al. (2016) assert that after accounting for inflation, population increase and increase in wealth, the increasingly popular notion of an increasing trend in losses from disasters disappears. As population increase leads to more lives being exposed to hazards, increased wealth also ensures more possession is exposed to drought risk. Inflation, on the other hand, ensures more recent losses are reported in huge figures as compared to the past. It is, however, the case that the increase in incidences and losses has seen the proliferation of efforts at drought risk management.

The increase in efforts at drought monitoring has, in general, lead to a spike in the use of remote sensing technologies and the available vast datasets for drought monitoring. On the global scale, despite the investments in both drought monitoring and disaster reduction being on the ascendancy, there still exist prospects for improving on understanding, monitoring and prediction of droughts (Wood et al., 2015). Equally, there is a need for drought monitoring systems that take into account practicalities of areas of interest through the investigation of the actual impact on lives and livelihoods as a result of drought episodes. There is, therefore, need for ground-based validation of drought monitoring data initiatives (Bachmair et al., 2016).

The current drought monitoring efforts are thus characterized by overreliance on a single indicator/ index, specialization in single crop monitoring, non-integration of socio-economic data to quantify impacts and non-comparison of alternative sources of drought monitoring data. Most monitoring systems are reliant on single indices like the vegetation indicator- the Normalized Difference Vegetation Index (NDVI) or have a reliance on a similar group of indicators like those derived from meteorological weather station data. A majority of these monitoring systems are also too specialized with over-concentration on models for monitoring specific phases of droughts like impact on selected crops without the incorporation of ground-truthing based on actual drought impacts on the society. The data used in the current monitoring systems are mostly sourced from single repositories without evaluation being done for appropriateness of purpose and fit for specific scenarios being undertaken.

The future of Drought Early Warning Systems (DEWS) is, for example, documented by Enenkel et al. (2015). The ideal DEWS is documented to include the key approaches of integration of remote sensing and socio-economic data, the thresholding of the integrated indicators and calibration of the integrated data into a Decision Support Systems (DSS). The remote sensing data types that could be used within DEWS include precipitation data, temperature data, evapotranspiration data and vegetation data. Precipitation data could from both ground-based rain-gauges and satellite-derived and modelled sources like Climate Hazards Group InfraRed Precipitation with Station data (CHIRPS) and Tropical Applications of Meteorology using SATellite (TAMSAT). Vegetation Indices (VIs) are mostly based on the Normalized Difference Vegetation Index (NDVI). Possible sources for vegetation indices are Moderate Resolution Imaging Spectroradiometer (MODIS), Land Satellite (LANDSAT), Advanced Very High-Resolution Radiometer (AVHRR) and SPOT Vegetation among others (Liu, 2015).

1.3 Problem Statement

There is an increase in both the frequency of droughts and the resulting economic losses from droughts, especially in the context of the GHA and particularly in Kenya and the wider East Africa. This spike in drought-related losses has led to the focus on Drought Risk Management (DRM) systems whose key elements are drought risk identification, drought monitoring, drought preparedness, and drought mitigation.

While drought risk identification involves the agreement on the definition of droughts based on some objective parameters, drought risk monitoring is based on the establishment of appropriate drought early warning systems (EWS) that signal the advent, progression and even possible cessation of drought events. As advocated by Mariotti et al. (2013), *drought risk identification* and *drought early warning systems* are the starting points to sound drought risk management that can greatly reduce the severity of social and economic damage by droughts.

The current drought early warning systems (DEWS) are characterised by four key features: popularity in the use of a single index, exclusion of the effects of drought, delayed availability of monitoring products and the tendency to use single techniques in the prediction of future drought conditions.

The tendency to use a single index for drought monitoring with precipitation the most used as recommended by WMO (2012). The tendency to use the single index is despite the proliferation of multiple indexes as reviewed in Su et al. (2017) and in AghaKouchak et al. (2015) and the segmentation of drought into four phases and hence types: meteorological, hydrological, vegetation and socio-economic droughts as documented in Hao, Singh & Xia (2018) and UNOOSA (2015).

The second characterization of the current DEWS is the fact that the aspects of human livelihoods that define socio-economic droughts are rarely part of these DEWS with most systems purely reliant on remote sensing data without an element of ground-truthing like the case in The African Drought Monitor (Sheffield et al., 2008), Klisch & Atzberger (2016) and the Famine Early Warning System, FEWS NET (Brown et al., 2015) amongst many other DEWS. The incorporation of socio-economic data as a possible ground-truthing in DEWS is advocated for in a few studies including in Bachmair et al. (2016), Hao, Singh & Xia (2018), Enenkel et al. (2015), Jenkins (2012) and Massarutto et al. (2013). The study in Hao, Singh & Xia (2018) for example documents the popularity of prediction of drought signals but with less effort invested in the prediction of the effects of such droughts on the society. The lack of systems that predict effects of droughts is the same gap identified in Bachmair et al. (2016) that surveyed 33 DEWS experts that advocate the inclusion of effects of droughts on the society as part of drought monitoring.

The third characterization of the state of art DEWS is that the vast implementations of the DEWS are ex-post or at best near real time (NRT). Existing DEWS thus mostly provide information at or after the lapse of the periods of monitoring. Moreover, the realisation of NRT systems is limited by the processing latencies that are inherent in the satellite-based data on which most of these systems are premised.

The final characterization of existing DEWS is that in addition to the above limitations, even instances of ex-ante systems mostly follow the common approach of searching for the single best performer/champion model, often using a single modelling technique. The use of the single index in a single technique is documented in a majority of studies with those in Ali et al. (2017) and Khadr (2016) as examples. Generally, predictive systems realised from the approach of selection of a champion model have a low predictive capacity that wanes in the prediction of future conditions as empirically proven in Meade & Islam (1998). Even in cases where multiple techniques are used, the objective is majorly the choice between competing model building techniques rather than the realization of synergies from the independent techniques.

From the characterization of the drought monitoring problem and the state of the art, this research focuses on gaps in the two perspectives to the drought prediction problem: the data perspective and the modelling process perspective.

1) **The data perspective** that is replete with data issues and the need for use of multiple indices covering the entire drought spectrum. With droughts having different definitions as documented in Lloyd-Hughes (2014), the use of

multiple indices in drought monitoring and drought prediction is a possible mitigation to the over-reliance on single-index models. The use of multiple indices makes sense much more in the face of the availability of many drought indices appropriate for drought monitoring across the different types of drought. Multivariate systems demand the identification and processing of the multiple datasets which is itself not a trivial undertaking (Bunting, 2017). The assessment of these datasets for appropriateness of purpose and their conversion to indices responsive to drought would thus be a logical step before their use for drought monitoring or drought prediction.

2) **Modelling process perspective** that should not only see the use of multiple indices across the different types of drought but also aim for highly predictive models through the harnessing of the different strengths of multiple prediction techniques. Such predictive systems realized from the combination of multiple techniques have been advocated to have high predictive performance that remains stable into the future (Hagedorn, Doblas-Reyes & Palmer, 2005).

The objective of the research is to build the ideal future drought prediction system with high predictive performance and future stability that also integrates multiple indexes from all the types of drought including data on effects of drought on society. The higher predictive performance and stability of the predictive system will be realized through model ensembling ideally built using multiple drought prediction techniques.

1.4 Research Objectives

1.4.1 Overall Objective

The overall objective of this research is to build and evaluate the performance of both homogeneous and heterogeneous models in the prediction of drought severity and drought effects using remote sensing and socio-economic data.

1.4.2 Specific Objectives

i. Determine the different biophysical and socio-economic variables that are used in the monitoring of drought and investigate their relationship with drought.
ii. Build and evaluate the performance of multiple drought prediction models using Artificial Neural Networks (ANN) and Support Vector Regression (SVR) as the case study Machine Learning methods.
iii. Build and evaluate the performance of homogeneous and heterogeneous ensemble models of both ANN and SVR in the prediction of drought severity and drought effects.

1.5 Research Questions

The research questions are mapped to the research objectives as provided in Table 1.

Table 1: Mapping of research objectives to research questions

Obj No.	Objective	RQ No.	Research Question
O1	Determine the different biophysical and socio-economic variables that are used in the monitoring/ prediction of drought and investigate their relationship with drought.	RQ1	What are the different biophysical and socio-economic variables that are used in the monitoring/ prediction of drought?
		RQ2	How do the variables identified for drought monitoring relate with drought?
O2	Build and evaluate the performance of multiple models for drought prediction using Artificial Neural Networks (ANN) and Support Vector Regression (SVR) as the case study Machine Learning methods	RQ3	What are the multiple models of both Artificial Neural Networks (ANN) and Support Vector Regression (SVR) that can be built for the prediction of both drought severity and drought effects?
		RQ4	What is the performance of the ANN models as compared to SVR models in the prediction of drought severity?
		RQ5	What is the performance of the ANN models as compared to SVR models in the prediction of drought effects?
O3	Build and evaluate the performance of homogeneous and heterogeneous ensemble models of both ANN and SVR in the prediction of drought severity and drought effects	RQ6	What is the performance of the Artificial Neural Networks (ANN) and Support Vector Regression (SVR) homogeneous ensemble models in the prediction of both drought severity and drought effects?
		RQ7	What is the performance of the ANN and SVR heterogeneous ensemble models in the prediction of drought severity and drought effects?

From Table 1, the three research objectives are considered achieved when the research questions are answered for each. The first objective is a function of both literature review and data preliminary analysis at the pre-modelling stage. The second objective is achieved by building multiple ANN and SVR models and choosing which are considered predictive of drought. Finally, the third objective considers the building of homogeneous and heterogeneous model ensembles using different methods and evaluating their performance with the traditional champion model approach as the baseline.

1.6 Significance

The study sets out to settle the comparative performance between heterogeneous and homogeneous model ensembles built using both biophysical and socio-economic data in the prediction of future drought conditions. The aim is to find out if model ensembling offers better returns in the prediction of future drought conditions as compared to the tradition single best model selection approach. The benefits of the study can thus be reviewed in terms of significance to the wider society and subsequently to both the research community and the practitioners in drought monitoring.

The definition of drought in this study is done in terms of vegetation conditions that is in itself a proxy to agricultural drought. On the other hand, the definition of drought effects is done in terms of malnutrition conditions for children under five years. Vegetation conditions closely mirror pasture and browse conditions. The communities in the study area having their economies mainly driven by pastoralism will relate to the results of the predictive system developed from this study. Proactive drought monitoring is bound to ensure minimized losses of both lives and livelihoods as a result of well-targeted drought interventions that are a product of better-formulated drought response plans. The collection of household data on drought effects will make for a ground-level driven monitoring system and given the fact that both predicted quantities are measured will assure the society on the objectivity of the predictions.

The government will have an opportunity to retune the current drought policies to have drought prediction as a minimal requirement of the drought early warning systems (DEWS). Such an opportunity will see the incorporation of policy elements that further enhance the “no regrets approach” to drought response and hence the possibility of incorporating intervention funding based on forecasts.

The practitioners and research community in drought monitoring will, however, be the greatest beneficiaries of this study. Since the study widely investigates the performance of ensemble models as compared to the current approach of single champion models and empirically grounds the superiority of stacked model ensembles, the research community has an opportunity to develop better predictive models that remain stable into the future. The study offers three approaches to model ensembling with a set of metrics to help choose which approach offers the most predictable models. The fact that the study also builds a model that down-scales spatially makes for an optimal approach for drought prediction for multiple spatial units.

The automation of model building is perhaps the most useful benefit to practitioners. With the reduction of human intervention in arriving at the models, objectivity is amplified and this widens acceptance of the model outputs within the wider research community. An additional benefit to both researchers and practitioners will be the pre-processed datasets that arise from this study. Even for cases out of the study area used for this study, the set of scripts can be shared for download and pre-processing of the data.

The prediction of socio-economic conditions as a result of future drought conditions will make for a set of directly actionable outputs out of a drought monitoring system that is expected to guide both drought preparedness and drought response in a model that can support feedback to the communities on terms and concepts commonly understood between them and the practitioners.

1.7 Assumptions

The following are the assumptions that will be made during the study.

i. The chosen data sources for the study were assumed to be available in the future for any efforts to replicate, extend and/or validate the results of the study.

ii. Transformation of data and selection of variables was undertaken as a part of the study and is premised on such transformations yielding variables useful in drought monitoring and the development of predictive models. Some of the transformations, to avoid data loss, are done from the data collection point at the pixel level.

iii. Although we assume that the data sourced from the operational data warehouses are representative of reality, we carried out an independent test of reasonableness on the data using statistical analysis methods.

iv. That the chosen area of study, being drought-prone, will continue to exhibit this tendency in the future and that the results being objectively obtained and documented lend themselves to generalizability beyond the area of study and into the future.

1.8 Scope

The main focus of this study was the need for predictive drought models with high predictive power and that integrate variables on the effects of droughts. The predictivity of the models is proposed to be achieved through model ensembles on data covering the selected study area for the period 2001-2017.

The research, therefore, focused on the identification of variables for drought monitoring, the identification of open access sources of the data, the extraction and pre-processing of the data. Subsequently, the formulation of ensemble models and the evaluation of their performance in the prediction of both drought severity and drought effects were undertaken. We addressed issues of variable selection between competing datasets, the over-production of models and subsequent selection of model ensemble membership and the incorporation of both remote sensing and socio-economic data.

The development of the experimental and investigative tools followed a review of literature and past studies and hence was grounded in theory. Two sets of scripts were developed. The first, Multisensor Remote Sensing Data Pre-processing (MSRSDP) tool, was developed as a series of scripts that automated the download and pre-processing of different remote sensing data. The scrips automatically download, spatially sub-set, smooth the data and correct for noise in the data and finally (dis)aggregates the data to a monthly frequency for the experimental phase of the study. The second set of scripts were for the development of the individual ANN and SVR models and their subsequent ensembling using multiple methods.

The modelling tool is scoped to be able to handle multi-sensor data modelling using multiple approaches. Such proposed approaches include the comparison of distributions, comparison of correlations and the comparison of seasonally of the data using seasonally adjusted correlations. For the selection between competing but duplicative datasets like TAMSAT and CHIRPS, the multiple metrics used included the spearman's correlation coefficient, Akaike information criterion (AIC), the relative importance of variables as partitioned by R^2 and the use of modelling approaches like support vector regression (SVR) and general additive models (GAM).

The modelling methodology used for this study integrated an ensemble of ANN and SVR techniques and investigated the performance of the different methods of the combination of multiple models in the model ensembles.

The study method did not model on multiple vegetation sensor data since only one NDVI remote sensing data source was included, together with rainfall data that is a choice between TAMSAT and CHIRPS. Other datasets included those that influence hydrological droughts like land surface temperature and evapotranspiration. We recognize that this approach leaves out other competing sources of remote sensing data.

1.9 Thesis Overview

The rest of the thesis is organised into chapters as follows: -

- *Chapter 2* aims to domicile the twin problems of drought prediction and the use of model ensembles in the realization of better predictive models. In this chapter, we define drought as an objectively measurable concept based on the key concepts of deficiency in precipitation, deviation from historical conditions, occurrence in space and time, intensity and duration and the idea of progression in drought conditions. The section also reviews drought monitoring systems that are both ex-post and ex-ante in nature to identify the remote sensing indicators that have been used to study the drought phenomenon. The basis for use of both remote sensing and socio-economic data in drought monitoring is established and a survey of both statistical and machine learning approaches carried out. The documentation, in literature of common methods used to realise highly predictive drought models is reviewed together with the accompanying algorithms. The ultimate objective of the chapter is the identification of similar works, gaps and the attendant possible methods that can be used to make contributions towards the realization of highly predictive models.

- *Chapter 3* documents the methodology used in the realization of the objectives of this research undertaking. This covers the identification and documentation of the data sources for the different kind of data required for this study and the subsequent methods of data collection for such data. The appropriate methods for the analysis as established with the tools and techniques outlined. The chapter presents the results from the pre-study that was run to establish the viability of both the methodology and the assumptions made for the study.
- *Chapter 4* presents the results from the different models developed and evaluated in the process of the study. The results are presented and discussed to make it clear on how the research questions are answered. The documentation follows on the order in which the objectives and research questions were formulated.
- *Chapter 5* outlines the summary of the major findings and contributions of the research into the integration of socio-economic data in drought prediction models and the performance of model ensembles in predictive drought monitoring. This section documents the limitations of the research, opportunities for practitioners and highlights the possible points for future research.
- Appendices that further support the results of this study are also provided. Such includes the full list of models developed and the validation for correctness of some of the assumptions made by the study.

Chapter 2: LITERATURE REVIEW

2.1 Prelude

The chapter undertakes to identify relevant literature, evaluate the sources and identify the gaps especially as relates to the prediction of droughts. The section aims to identify what has been and what has not been investigated, outline the key datasets used, review how key concepts have been defined and how these concepts have been ultimately measured. Besides, the section provides evidence on which basis the findings of the study are supported. Ultimately, this section offers theoretical underpinning for this research. In it, we define the problem of drought and drought prediction as a specific type from the general set of prediction problems that are premised on using specific past examples to generate a broad generalization of the future. A targeted review of past studies and the trends in drought monitoring is presented in this section.

2.2 Definition of drought

Drought is one of the natural disasters that is the most widespread and strongest felt even though it is not widely understood due to its causes being as a result of the interaction of multiple complex factors.

Drought is defined in diverse ways, the common of which is the deficiency in precipitation over an extended period, usually a season or more, resulting in the shortage of water causing adverse impacts on vegetation, animals and/ or people and thus hindering various economic sectors like agriculture, industry, hydropower generation and recreation sectors (Bordi et al, 2005; Morid, Smakhtin & Bagherzadeh, 2007; Schipper, 2003; UNISDR, 2009). It is noted, however, that drought does not have a direct one sums it all definition and interpretation as earlier documented in Palmer (1965) and even recently in Lloyd-Hughes (2014). The view of deficiency of precipitation, based on the definition of drought above will, therefore, be sector and time-specific and based on some concept of anomaly/ deviation from some expected conditions.

The other aspects noted in the definition of drought include its characterization as the most complex natural disaster that is however less understood as documented in Morid, Smakhtin & Bagherzadeh (2007) and Ali et al. (2017). That drought is a complex natural disaster is attributed to both the difficulty in definition their beginning (onset) and/or end (off-set) and their tendency to also lead to and be accompanied by other disasters like extreme heats and wildfires.

As opposed to other disasters, droughts are described as slow on-set hazards and hence viewed as a creeping phenomenon with which comes the benefit of time that could be used to undertake effective mitigation and preparedness measures (UNISDR, 2009; Wilhite, 2006). Droughts are documented to exhibit a rarity in occurrence as compared to other natural disasters as illustrated in Figure 2.1. For example, from Figure 2.1 it is clear that drought is indicated to trail other distinct disasters like earthquakes and floods in the frequency of occurrence by a factor of at least 5.6 (EMDAT, 2012)

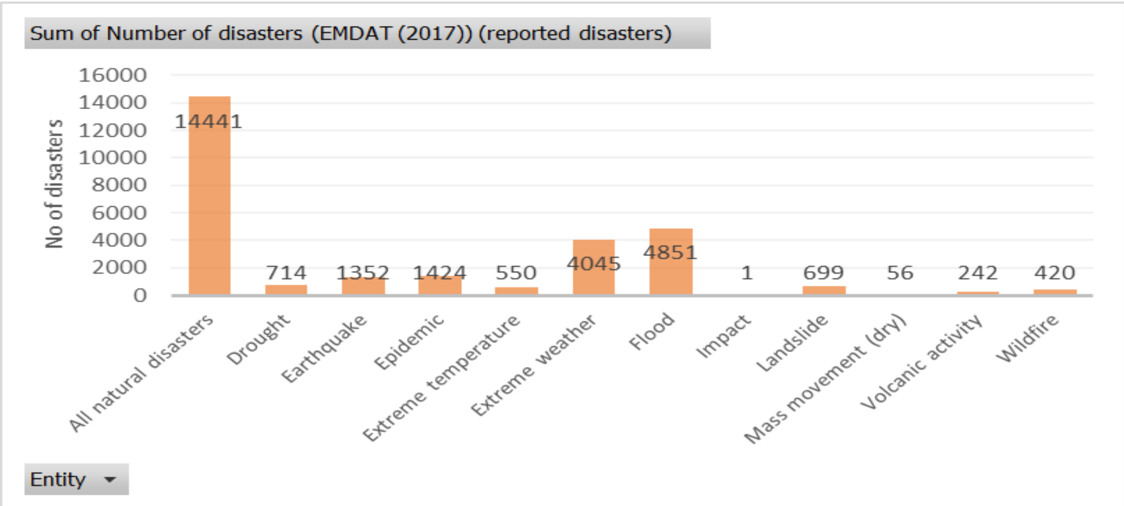


Figure 2.1: The frequency of different world disasters (1900-2017).

As evidenced in Figure 2.2, over the decades from the 1900s and despite the rarity in the occurrence of drought as compared to other major disasters, the social impact of drought as indicated by the number of deaths supersede those of other disasters. This observation was made earlier in (Hewitt, 1997). This is perhaps a direct result of the tendency of drought to cover wider areas and thus guaranteeing greater impacts on both lives and livelihoods.

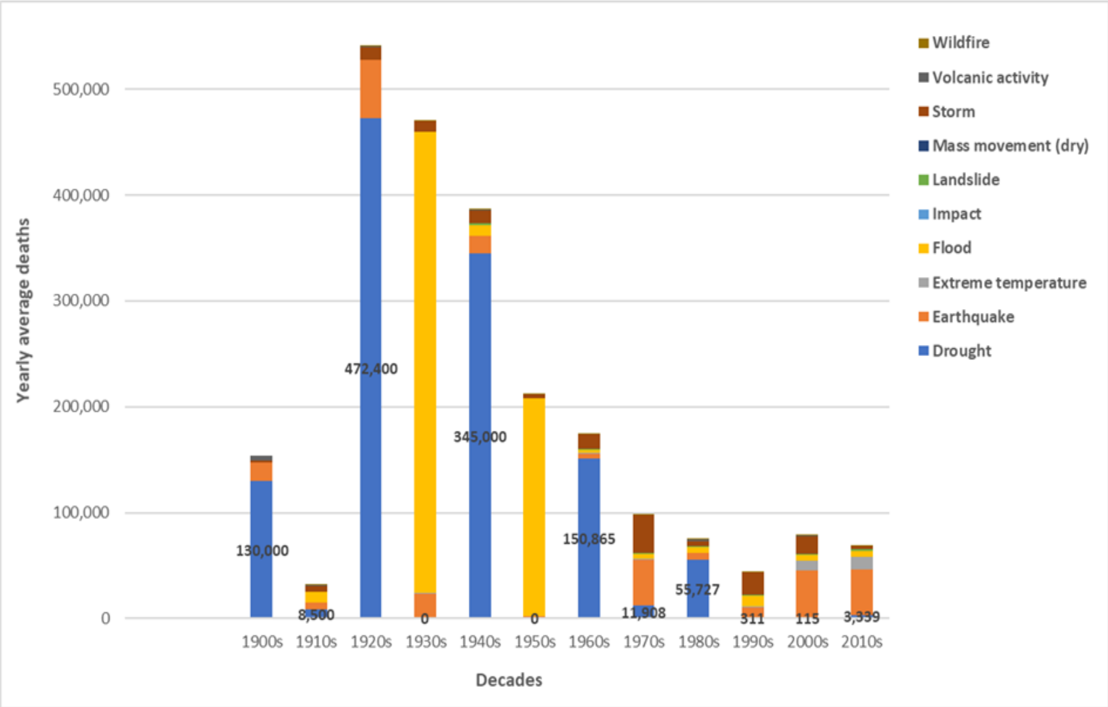
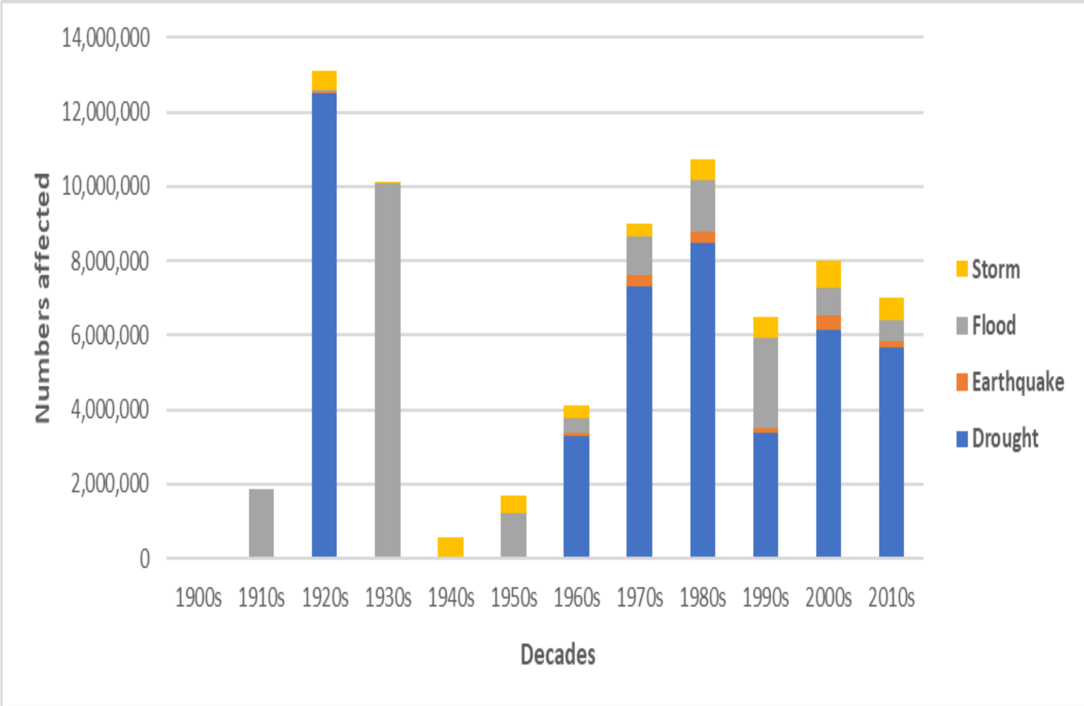
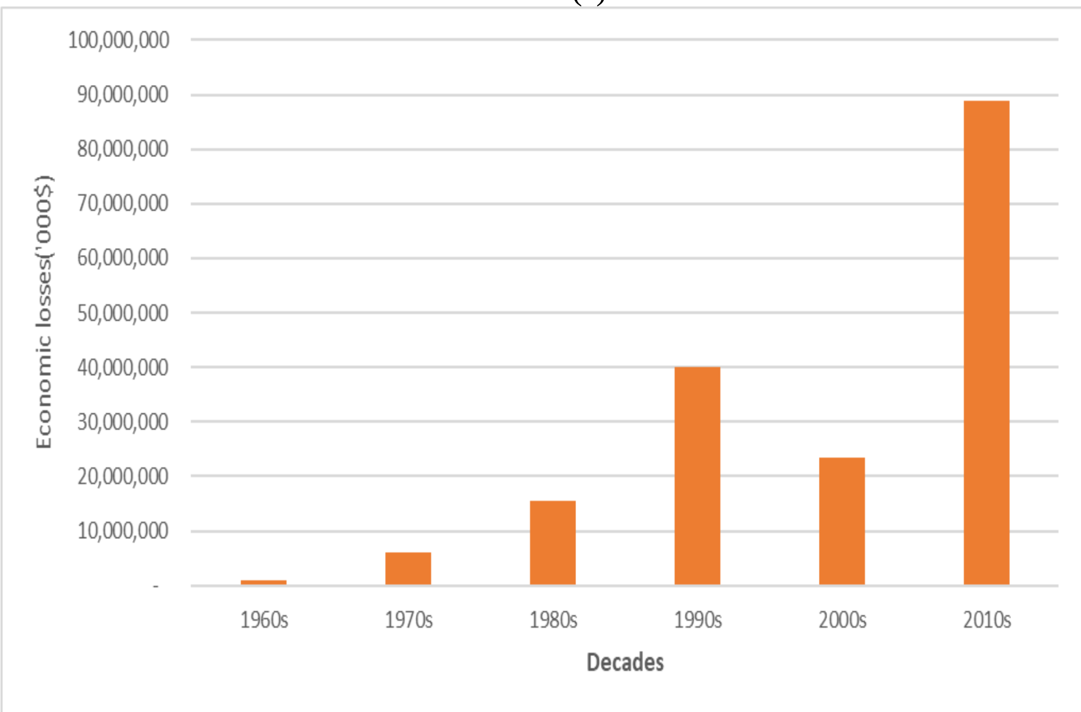


Figure 2.2: Global average deaths from selected natural disasters by decade for the decades 1900s-2010s (EMDAT, 2012).

Despite the documentation of the reduction in the average number of deaths attributable to drought over the decades (EMDAT, 2012), it is also the case that the numbers affected by drought as shown in Figure 2.3(a) is higher than that for earthquakes, floods and storms. Furthermore, the economic losses due to droughts have been rising over the years as shown in Figure 2.3(b).



(a)



(b)

Figure 2.3: Number of people affected by different disasters (a) and the economic losses from drought over the years-1965-2019 (b) (EMDAT, 2012).

Zeroed in to Kenya and East Africa, there is an increase in both the frequency of drought and the cost of economic losses as a result of droughts particularly in Kenya and East Africa. For example, Government of Kenya (2012) documents the 2008-2011 drought in Kenya as having made 3.7 million people food insecure with economic losses approximated at US$ 12.1 billion. These losses in Kenya mirror the global scenario where the 2014 California drought was projected to have cost a total of US$ 2.2 billion in losses as documented in Cody, Folger & Brougher (2009) that also observe the increase in the ability of droughts to cause widespread misery. Further documentation of losses due to drought are reviewed in Ding, Hayes & Widhalm (2011) with a caution on the use of the estimates due to differences in methodologies and the non-documentation of some droughts with localized impacts given limited spatial and temporal extents of coverage.

2.3 Key concepts in the definition of drought

Despite the non -existence of a uniform definition of drought as a result of differences in its perception, there are a set of concepts common to most definitions. For the investigation of the drought phenomenon, given this very non-universality in its definition, there is a need for the identification and documentation of these key concepts and characteristics. In fact, as documented earlier in Wilhite (1993) and recently in Lloyd-Hughes (2014), the absence of a universal definition of drought is perhaps the source of differences in the methodologies for drought monitoring.

The absence of a universal definition of drought is, however, not a problem in itself. This is because drought affects multiple sectors of an economy and livelihoods and thus perspectives are bound to differ in its definition. Despite the differences in the definition of drought that could be either conceptual or operational, the following key concepts keep recurring: -

- ***Deficiency of precipitation***

  The greatest basis of the definition of drought and its sub-sequent monitoring is the deficiency of precipitation over a period of time. This is a pointer to precipitation as the key variable for drought monitoring. This definition is adopted by many studies including in Bordi et al. (2005) and Morid, Smakhtin & Bagherzadeh (2007) and those provided in the earlier section on drought definition.

- ***Deviation from some historical conditions***

  The deviations in precipitation, even when drought is defined in terms sufficiency of precipitation has to be considered abnormal. The deviation is a measure based on some normal conditions that are often defined based on an identified historical period. The abnormal deviations are referred to as an **anomaly**. The concept of anomalies is for example documented Hayes et al. (2011) in the review of precipitation deficits compared to the historical averages for a given region. Drought monitoring would, therefore, be equated to the process of monitoring the occurrence of these anomalies in the precipitation and indeed of any other drought indicator over a given spatial extent. A comprehensive review of some of the anomalies is provided in Eslamian et al. (2017) and Zargar et al. (2011).

- ***The occurrence in space***

  The definition and occurrence of drought are both based on space. The concept of space, in terms of drought monitoring, implies a location, area or polygon. This implies that what is drought in one location should not necessarily be a drought in another location. Augured with the concept of anomalies, this implies that droughts, their occurrence and severity are defined by occurrence in a spatial extent. The definition of drought with the concept of spatial extent is almost universal and is for example in Adede et al. (2019a), WMO (2012), Klisch & Atzberger (2016), Klisch, Atzberger & Luminari (2015) amongst many other studies. The concept of space in the definition of droughts goes together with that of time thereby implying droughts as having both a spatial extent and temporal extent.

- ***The occurrence in time***

  The definition of normal conditions for a spatial extent or location is based on some historical reference. Also, all spatial extents more or less are expected to have periods of reduced precipitation. Not all these periods should be interoperated to be drought episodes. Drought periods must, therefore, ideally, also be defined for the same time in history. The definition of the popular indices like standardized precipitation index (SPI) in WMO (2012) and vegetation condition index (VCI) in Kogan (1990) are quite strong in applying the concept of time in the definition of drought. The concept of time, like of space defined above, are also used in Adede et al. (2019a) and in Adede et al. (2019a) together with Klisch & Atzberger (2016).

- ***Severity and duration***

  Once established, the severity of a drought is a measure or quantification of its deviation from a set threshold as defined by the run theory in Zhang et al. (2015). Duration, on the other hand, is mostly referred to as the period of occurrence of a specific drought and involves the capability to pinpoint the period between the onset and the off-set of the drought.

- ***The idea of slow onset and progression in time and space***

  Droughts are considered slow onset with the ability to progress both in space and time (UNISDR, 2009; Wilhite, 2006). This slow onset nature of droughts is the characteristic that lends droughts towards being a monitorable and predictable event that offers time for intervention planning. The slow onset view of droughts has however been challenged by Basara et al. (2019) and Otkin et al. (2016) that documents the 2012 Continental US flash drought that was modelled using variables like evapotranspiration, soil moisture and vegetation conditions in addition to precipitation.

- ***Effects on sectors***

  Droughts, as documented variously and specifically in Bordi et al. (2005), Morid et al. (2007), Schipper (2003) and UNISDR (2009) are documented to have effects or impacts on different sectors of the economy. While some effects are direct, others are indirect and affect the socio-economic aspects of societies.

2.4 Types of drought

Apart from the identification of the key concepts in the definition of drought, the alternative view to a global definition is the classification/categorization of droughts into types. The alternative methods in the classification of droughts are documented in Demuth & Stahl (2001), Monacelli, Galluccio, & Abbafatim (2005) as based on either of the basis on which the drought is defined or on the discipline of practice.

- ***Formulation of drought as the basis for classification***

  Based on the formulation and definition of droughts, they can be classified into two distinct types as either conceptual or operational. Conceptual droughts have no basis for rolling assessments and are thus broad and general in their definition. On the other hand, operational droughts have the determination of onset, severity and offset of drought episodes as their objective. Operational drought monitoring is therefore based on some concept of an operational definition of drought. This approach makes the study of droughts a more structured concept since we can define it in terms of on-set, off-set, duration, severity and even extents of occurrence.
- ***The disciplinary perspective of drought as the basis for classification***

  Disciplinary perspectives to drought definition use the concepts of operational drought to define droughts as viewed by different sectors of practice. This approach gives the categories of drought as any of meteorological, hydrological, agricultural, and socio-economic droughts. The description of the droughts based on their types by discipline is provided in UNOOSA (2015) and is amplified in drought monitoring studies. These different types of drought are elaborated as follows:

  - Meteorological drought is defined in terms of the magnitude of a precipitation shortfall and the duration of this shortfall event and is noted to be a broader definition of drought as opposed to the other types of droughts. Studies that document meteorological droughts include Bordi et al. (2005), Huang et al. (2016), Khadr (2016).

- Hydrological drought is usually based on surface and sub-surface water supplies. It is always closely related to agricultural drought. The indexes used to monitor this type of drought are therefore based on water sources. An example study that document hydrological drought is, for example, Kubiak-Wójcicka & Bak (2018) that is based on river flow.
- Agricultural drought is mostly defined in terms of the impacts of droughts on agriculture as indicated by precipitation deficits, soil moisture and evapotranspiration. The monitoring of agricultural drought is for example in Gu et al. (2008), Klisch & Atzberger (2016), Svoboda et al. (2002) and Tadesse et al. (2014).
- Socio-Economic drought is the result of droughts affecting people, lives and livelihoods. It is therefore characterized by the destabilization of normal demand and supply systems of some economic goods with the progression of a drought event. Possible use of socio-economic data in drought monitoring is provided in Enenkel et al. (2015), Jenkins (2012) and Garrido (2014).

Adopting the operational definition of droughts that covers the entire of the types of disciplinary drought definition as advocated for in Enenkel et al. (2015) would be a powerful formulation of an operational drought monitoring system. A sound DEWS would, therefore, be indicated by the objective definition of drought that incorporates the key characteristics of drought identified in section 2.3 that includes: spatial extent, temporal coverage, severity, duration and the definition of reference/normal conditions and periods.

2.5 Drought Risk Management (DRM)

The quantification of drought, just like its direct modelling, remains a difficult undertaking. This is as opposed to the effects of drought that are, however, identifiable and are perhaps measurable. The management of droughts can thus be viewed as the minimization of drought risk. Drought risk can be viewed as the likelihood that drought leads to casualties, damage or loss.

With the risk view in mind, drought can be viewed as a natural hazard whose occurrence leads to a disaster. It is, however, the case that the drought hazard has a chance or probability of occurrence when defined in the confines of space and time. This probability of drought occurring is defined as the risk to drought and is what the drought risk management initiatives centre around.

There are two approaches to handling the problem of droughts. First, is the episodic approach that reacts to droughts at their time of occurrence. Second is a strategic approach that adopts the definition of drought as a risk and proactively manages the risk of droughts. Given that droughts are natural disasters, the choice between these approaches then begs the question on the methods for drought risk management.

The traditional approach to drought management, in which droughts are monitored and responded to as and after they unravel, and the learning on the losses from past drought events, has led to a shift towards Drought Risk Management (DRM). DRM is considered a holistic approach that includes the following: -

- ***Drought sensitive policy formulation***

  This step involves the definition of drought-sensitive policies at global, regional, national and even sub-national levels that are geared towards ensuring droughts do not lead to losses. The formulation of global bodies with mandates on drought monitoring is one of the initiatives at this level of drought risk management (DRM). In Kenya for example, there has been the formulation of the Ending Drought Emergencies (EDE) under the Vision 2030 (Vision 2030, 2007). The effect of the formulation of drought-sensitive policy is to mainstream drought risk management in the normal programming of any given country.

- ***Prioritization of drought monitoring***

  Drought losses have led to the realization of need by governments, including GoK, on the need to have an effective Drought Early Warning System (DEWS). DEWS provide information on occurring hazards that might evolve into disasters unless early response and possible mitigation measures are initiated. The objective of a DEWS is to monitor droughts in an objective way that also tracks the evolution of the hazard through stages that are part of an overall management cycle. The DEWS implemented for drought monitoring should be comprehensive and responsive to environmental and climatic events. ILRI (2009) notes that Drought is the prime recurrent natural disaster in Kenya that affects up to 10 million people, mostly pastoralists. ILRI notes that, despite a National Drought Management System being in force in Kenya for almost 20-year, one common limitation that is highlighted is that the systems remained static in the past with methodologies in place long after they were established. As documented by ILRI (2009), the Government of Kenya (GoK) and the European Union (EU) initiated a process to review both the DEWS and the drought response strategies in Kenya. This particular initiative aimed to make an evidence-based drought early warning system that together with assessment of vulnerabilities can form a basis for early response to drought. The evidence-based system is expected to ride on the fact that, in the context of Kenya, drought is a slow onset hazard (UNISDR, 2009) that by and large provides a window for response and mitigation.

- ***Legislation on and Institutionalization of drought management initiatives***

At the national level for many governments, there has been witnessed the establishment of specialized institutions for drought and drought risk management. The National Drought Management Authority (NDMA) in 2011 and subsequent legislation of the NMDA in Act No 4 of 2016 as a permanent drought management body is one of the successes of the influence on policy that is attributed to drought and the effects of droughts. The NDMA is in charge of drought monitoring and resilience building with the overall aim of ensuring droughts do not become disasters. Other initiatives include the establishment of a Drought Contingency Fund project (DCF-P) that provides funding for both drought preparedness and drought response activities and its institutionalization through the establishment of the National Drought Emergency Fund (NDEF).

- ***The adoption and implementation of a Drought Risk Management (DRM) Framework***

Drought risk management (DRM) is mostly driven by frameworks that have been formulated over the years. One such framework is the Hyogo Framework of Action, HFA (UNISDR, 2005) while another is the UNISDR principles to drought risks management. These are summarized as shown in Figure 2.4.

No	Hyogo Framework for Action	UNISDR Principles
1	Make Disaster Risk Reduction a Priority	Policy and Governance
2	**Know the Risks and Act**	**Drought Risk Identification, risk Monitoring and Early Warning**
3	Build Understanding and Awareness	Drought Awareness, Knowledge Management and Education
4	Reduce Risks	Reducing underlying factors of drought risk
5	Be Prepared and Ready to Act	Enhancing mitigation measures and preparedness for drought

Figure 2.4: Hyogo Framework for Action (HFA) Priorities for Action (PFA) and UNISDR guiding Principles from the Sendai Framework

At the very basic, the risks of droughts can be modelled as shown in Equation 1& 2. The approach of risk management is to view drought risk as a function of the elements of hazard (H), vulnerability (V), exposure (E) and capacity (C).

$$R = \frac{(H \times V)}{C} \quad \text{……………………………………………………………….. (1)}$$

$$R = H \times E \times V \quad \text{…………………………………………………….……….. (2)}$$

Equations 1 and 2 are basically the same since both assert that drought risk (R) directly depends on the hazard (drought, H) and vulnerability (V) for the exposed elements. In equation 1, drought risk (R) is mitigated by the capacity (C) of the exposed elements. On the other hand, equation 2 assumes vulnerability (V) as what remains after capacity (C) is accounted for.

The above Equations are a reduction of the Drought Risk Management (DRM) framework that is an extracted from the Hyogo Framework of Action, HFA and the Sendai Framework's UNISDR principles (UNISDR, 2005).

The key elements of drought risk management (DRM) therefore include: drought contingency planning, drought early warning, drought resilience building, drought preparedness, drought impact assessment, drought communication, drought response and drought recovery all carried out in the context of a drought early warning system (DEWS). Absence of these aspects of DRM, especially that of DEWS has resulted in reactive systems as opposed to the demand for proactive systems. One key tenet of the DRM approach is the need for reliable prediction-driven systems and models that are incorporated as part of the DEWS. In fact, it is our observation that a predictive drought early warning system that provides a preview of the expected future condition is the one key weakness that plagues most of the existing drought early warning initiatives.

DEWS are by their nature domiciles in the second action point of the Hyogo framework (Know the risks and act) as well the UNISDR principle on Drought Risk Identification, Risk Monitoring and Early Warning systems. The early provision of drought information for early action would thus be viewed as a key undertaking of DEWS.

2.6 Drought Monitoring Systems

From the key concepts of drought in section 2.3, four concepts are particularly important when it comes to operational drought monitoring. As also defined in section 2.4, operational drought monitoring is concerned with the definition of drought severity, onset and offset. The four concepts of drought deemed important are: droughts having both a spatial extent of coverage and temporal coverage (Tsakiris et al., 2013); droughts as slow-onset disasters (UNISDR, 2009; Wilhite, 2006) and droughts as being progressive along some objective measure of severity (UNOOSA, 2015). The implication is such that, we can monitor drought for a given location and time with the benefit of time to quantify the changes in its severity. With the benefit of time as drought progresses, it is then possible to model drought impacts for a spatial extent of interest. Such operational monitoring would be achieved through drought early warning systems (DEWS).

A DEWS would thus have its key goal as the operational monitoring of droughts in the context of their severity and progression in time over a given spatial extent under consideration. The establishment of a DEWS would thus be justifiable based on the realization that better improved and timely drought-related decisions lead to reduced impacts on both people's lives and livelihoods. Typically, a DEWS would have a selected extent of coverage with the possibility of having multiple statistics for multiple extents over a selected frequency of monitoring that could be any of daily, pentadal, weekly, decadal, bi-weekly, monthly, quarterly, semi-annually, annually etc.

To achieve operational drought monitoring, drought has to be defined. According to Wilhite (2006), Morid, Smakhtin & Bagherzadeh (2007) and Bordi et al. (2005) drought can be viewed as a cumulative departure from normal or expected conditions/levels of precipitation. The normal is defined differently in various monitoring systems. The most common definition of normal is the long-term average. The acceleration of drought and effects of droughts once onset is documented to be at varying speeds for various spatial extents. This, therefore, makes it necessary to have in-place drought monitoring systems to areas prone to drought related disasters.

Drought monitoring systems are thus deployed to identify changes in climate and hence aim to detect the likelihood of occurrence and the expected severity of drought. The key is to avail drought related information to decision makers in time to make drought risk management an active and continuous process rather than a reactive process. Drought monitoring systems have thus been used to provide historical records to assess changing conditions and thus provide early warnings of potential drought threats.

Due to the complex nature of drought as a disaster, Drought Early Warning Systems (DEWS) have been noted to be more complex as compared to those of other natural disasters (UNISDR, 2006). This complexity has ensured that they are less developed as compared to their peers, like say for floods. Ground sourced precipitation measurements and the related precipitation anomalies have been used as the common parameter for monitoring droughts. Recent developments include incorporation of Remote Sensing data in the form of rainfall estimates (RFE) and Vegetation Indexes (VIs) as proxies to drought monitoring (Niemeyer, 2008).

DEWS that are well designed should meet the three-point criteria guided by objectivity, sensitivity to small changes and communicability.

- ***Objectivity*** ensures that the definition of drought is well thought out and thus not subjective. In operational drought monitoring, once a definition is adopted, it is retained throughout the monitoring of the event. Even for cases where drought classes are used, they must be objectively determined. Objectivity builds confidence in a DEWS (Quiring, 2009) and is characterized by the ability to identify and use representative and reliable data. Any forecasts made should also be based on the integration of data from different sources including those on drought impacts possibly realized from field-based assessments.
- ***Sensitivity to changes*** as advanced in Wilhite & Svoboda (2000) ensures a DEWS has the ability to detect on-set and secession of droughts that are key milestones that determine the kind of responses required. DEWS must be able to signal drought situations before they occur even for localized cases of drought. The ability to detect these changes, however small is what sets apart DEWS in terms of their ability to contribute to the management of drought events especially when famines are ab outcome.
- ***Ease of and regularity in the communication*** of the products/outputs of a DEWS is a key requirement as documented in Grasso & Singh (2011). The systems generally remain non-technical in their outputs so as to be useful to a wider audience. Dissemination and communication of results stakeholders should also be in a timely manner. Objectivity in the definition of drought and consistency in the application of DEWS improved communicability as results are then trusted.

DEWS, based on a review of literature, can be categorized based on their extent of coverage and on their timelines for delivery of their products and outputs. Geographical coverage is used to popularly characterized the systems into ***Local (sub-national) systems*** that are majorly within small spatial extents, ***National systems*** whose coverage is a whole country, for example, the system in Klisch & Atzberger (2016) and the United States Drought Monitor (Svoboda et al., 2002), ***Regional systems*** that cover more than a country like the case of the African Drought Monitor (Sheffield et al., 2008) and ***Global systems*** whose coverage is the whole globe like the systems in Hao et al. (2014). It is generally the expectation that the larger the coverage, the more generalized the results and thus the more information loss occurs. Such large-scale DEWS are bound to miss the small extent and restricted time-scale occurrences of droughts.

On the other hand, timeliness in the delivery of products separates the systems into ***ex-post systems*** for example in Hayes et al. (1999) and in Brown et al. (2015); and ***ex-ante systems*** (Tadesse et al., 2014; Wardlow et al., 2012). Ex-post and ex-ante systems are also referred to as monitoring and predictive systems respectively.

Even though the terms Drought Monitoring System (DMS) and Drought Early Warning Systems (DEWS) are used interchangeably, it is our opinion that the difference should be that a DEWS offers a longer lead time as compared to a DMS in the provision of monitoring products and thus majorly driven by prediction of droughts. DEWS must be ex-ante systems while DMS should ideally be optionally ex-ante. With the ever recurring and intensifying reequipment for provision of early and timely information, DMS face the increasing need for predictive components.

2.7 Data and Indicators for Drought Monitoring

The process of operationalizing drought early warning systems (DEWS) involves the collection of data (Wilhite, Sivakumar & Wood, 2000) for the extents of interest. The data collected is basically of four categories: Precipitation data, hydrological data, vegetation data and in some cases, socio-economic data. A review of the data for drought monitoring is provided here following in the earlier classification of droughts based on operational monitoring systems into Meteorological, Hydrological, Agricultural and Socio-Economic drought. Remote sensing data are discussed in the light of Meteorological, Hydrological and Agricultural Drought while Socio-Economic data is done in the context of Socio-Economic drought.

Figure 2.5 provides a model for the discussion on the data requirements for drought monitoring. The model is based on UNOOSA (2015) that documents different types of drought and asserts their successive nature. The drought types of meteorological, hydrological and agricultural are noted to follow each other with the progression of the deficiency of precipitation.

Figure 2.5: Drought types and their progression.

The different types of drought are operationally defined differently. Meteorological drought is defined in terms of the deficiency of precipitation (degree of dryness) and the period of the deficiency, mostly over more than a season. Agricultural drought, on the other hand, is defined by the effect of the deficiency of water for plant growth and soil moisture. Hydrological drought is characterized by the drying up of surface and sub-surface water sources.

At the extreme end of a drought are the effects on the socio-economic indicators of areas of coverage including supply and demand for economic goods like water, milk, forage, food prices and hydroelectric power. Socio-Economic droughts are characterized by demand surpassing supply of the climate-dependent economic goods.

2.7.1 Remote Sensing data for drought monitoring

Remote Sensing is a data collection technique at the core of which is a sensing device that collects information from an object without physical contact. The greatest application of remote sensing would be to get data from instances where accessibility is a limitation (Patruno, Fitrzyk & Delgado Blasco, 2020). The principles of remote sensing of efficiency and effectiveness have been variously documented (Berhan et al., 2011; Few, 2009; Huete et al., 2002; Ojala, 2003). There is an increasing availability of remote sensing data; even at high resolutions with some of these sources having some level of pre-processing for the data.

The popularity of remote sensing data (RSD) for drought monitoring is attributed to the following reasons based on the desirable characteristics of the data as documented in Wardlow & Anderson (2012) and Gu et al. (2008): -

- ***Greater spatial and temporal coverage*** is provided by Remote Sensing of drought conditions than from site measurements of precipitation, soil moisture (SM), Land Surface Temperatures (LST) and vegetation cover. The sensors, having an automated approach are thus able to cover larger regions and at ever-increasingly high resolutions.

- ***Timely provision of information*** from remote sensing through variously automated data acquisition processes ensures non-delay of operational processes. The data satellites deliver data at regular intervals and repositories have the data provided on time at frequencies of design and contract. Planning around this attribute is made easy as products end up having timely delivery.

- ***Non-biased and accurate information*** out of remote sensing processes ensures reliability and trust in the data. Data accuracy issues are mostly due to factors like cloud cover and atmospheric interference that have methodologies for their corrections. As opposed to human-driven data collection, satellites are setup for objectivity, especially after design limitations are accounted for.

- ***Spatially continuous measurements*** provided by remote sensing data ensure data is available over large geographical areas including areas where access is limited and ground observations sparse or virtually non-existent. This is as opposed to other methods that have discrete and possibly categorized data collected over some desired extent.

- ***Consistent frequency of revisit*** of the satellites ensures that the data is regular and can thus be used to form trends for the same spatial coverage over the same period in history. This is perhaps the property that makes remote sensing data appropriate for operational drought early warning systems.

- ***Availability of historical data*** from remote sensing data repositories ensures the possibility of comparisons based on the historical cases for the detection of anomalies, which is the best way to define droughts. The historical data helps to support the building of historical trends capable of delivering value to studies that model changes over time. The availability of historical datasets ensures the building of models can befit both from being undertaken immediately and being validated prior to deployment for operational use.
- ***Simplicity in the calculation of anomalies indices*** and makes for the popularity of the use of remote sensing data. The indices also lend themselves to data visualization using different techniques like maps, dashboards, tables, charts, and matrices. The interpretation of the indices at times, however, becomes a limitation requiring some level of skill.

From the studies in Wardlow & Anderson (2012) and Gu et al. (2008), remote sensing applications for drought monitoring, therefore, require data sets that, first, hold the ideal characteristics of being able to be incorporated into operational data production that is routine (regular intervals- dekadal, monthly, 3-monthly etc.), The second desirable characteristic is the possession of a historical archive that can facilitate the calculation of anomalies based on any of but not limited to the per cent of average, relative ranges and standard deviations. The third is that the data be highly available and easily accessible, including availability in multiple formats that are relatively inexpensive to extract from and deploy into monitoring systems. The fourth desirable characteristic is that the data should be amenable to validation and reproducibility by subject matter experts, especially across multiple locations and time periods. Additionally, the data for drought monitoring should be location-aware and also be location sensitive for better identification of possible hotspots as advanced in Few (2009).

Key issues in the selection of remote sensing images for any purpose is, apart from the intended application, guided by the following considerations:

- Repetition rate/ frequency of the sensor/ satellite
- The spectral resolution that defines the number of spectral bands of the electromagnetic spectrum covered by the satellite
- Spatial resolution that defines the unit area for which data is collected. This unit is referred to as the pixel size.
- Cost as compared to the value offered by the information. In cases where open-source data is fit for purpose, cost is a justification enough to choose them.

Another element that is important but is often overlooked in the identification and use of remote sensing data is the aspect of ***data formats***. In the most. Remote sensing is sourced in the form of *raster images* in any of the valid formats including, but not limited to, TIF, GeoTIFF, ECW, GRID, IMG, JP2, SID. The images either come with the metadata as part of the images like is in GeoTIFF or a separate file as is in the case of IMG files. The choice of processing tools at times constrains the data formats to be preferred thought conversion from one format to another is widely supported.

2.7.2 Remote Sensing Indicators for drought monitoring

In this section, we review how remote sensing data is processed to realize drought-sensitive indicators. We then proceed to outline some of the important drought indicators and how they are realized from the raw datasets.

For the remote sensing data to be used for drought monitoring, the data must be processed into indicators that both quantify drought and that are at the same time sensitive to changes in drought conditions.

The definition of the indicators is both a data problem and an interpretation problem. A data problem in the sense that the indicators have to be based on some specific data that is guaranteed to be available in the future. The interpretation problem implies that the indicators will need to be easily interpreted, especially in operational drought monitoring. Since drought is widely defined as a deviation from some normal conditions, most of the data transformations are *"difference"* transformations between images of interest and some reference periods. A summary of these calculations/ transformations of the remote sensing data are as provided in Table 2.

Table 2: Common "Difference" indicators used in drought monitoring.
Most common is the definition of absolute, relative and standardized differences

Transformation	Transformation Formula
Absolute Difference to the historical median	$AD_{hm}(y,p) = X(y,p) - Median(p)$ ---(3)
Absolute Difference to the historical average	$AD_{ha}(y,p) = X(y,p) - Mean(p)$ ---(4)
Relative Difference to the historical median	$RD_{hm}(y,p) = [X(y,p) - Median(p)]/MEDIAN(p)$ ---(5)
Relative Difference to the historical mean	$RD_{ha}(y,p) = [X(y,p) - Mean(p)]/Mean(p)$ ---(6)
Standardized Difference	$SD_h(y,p) = [X(y,p) - Mean(p)]/StDEV(p)$ ---(7)
Relative Range Difference	$RR_h(y,p) = [X(y,p) - Min(p)]/[Max(p) - MIN(p)]$--(8)
Historical Probability	$HP_h(y,p) = Prob\ of[X(y,p)]\ in\ hist.distribution$---(9)
Historical Rank	$HP_h(y,p) = Rank\ of[X(y,p)]\ in\ hist.distribution$---(10)

Table 2 shows the difference transformations frequently used to realize drought indicators. The difference indicators popularly used include:

- The absolute difference indicators (Equation 3 & 4) which subtract one image from a reference image that is any of the mean or median images.
- The relative differences (Equation 5 & 6) that further calculate the ratio of the "*difference"* transformations to the reference mean or median image used in the transformation. In essence, the relative differences make the actual differences less pronounced.
- The standardised difference (Equation 7) and relative range difference (Equation 8) are perhaps the most popular. The standard difference approach gets the number of standardised deviations an image is away from the mean of a reference historical period. A variant of this is recommended by WMO (2012) in which the new image values were resampled such that the mean is zero (0) and the standard deviation is one (1). The relative range, as opposed to the standardised approach, transforms the difference between the current value and the minimum value from the reference period with the range of the values from

the reference period. In essence, it stretches the current value within the range if the minimum and maximum historical data mostly from a similar period in history. An evaluation of this stretching is for example used in Kogan (1990) and in Klisch & Atzberger (2016).

- The historical probability (Equation 9) and historical rank (Equation 10) that both calculate the probability and rank of the current values in the historical distributions respectively

The above transformations of remote sensing data depend on data and are in essence only useful if parameters to be monitored from space are identified and respective datasets made available in formats considered appropriate for the calculation of the indexes.

The possible parameters to be monitored from space using Remote Sensing technologies for the different phase of the drought are documented in (Khamala, 2017; UNOOSA, 2015; Zargar et al., 2011). The parameters to be monitored include, but are not limited to Precipitation (PPT), Surface Water Storage (SWS), Ground Water (GW), Land Surface Temperatures (LST), Evapotranspiration (EVT), Snow, Soil Moisture (SM) and Vegetation (VGT). The parameters are mostly monitored at Global, Regional, National and sub-national levels.

The remote sensing indicators, reviewed following on the types of droughts earlier defined are described as provided here next.

- ***Meteorological drought indicators***

Meteorological drought monitoring is mostly based on precipitation. Precipitation, in the forms of rain, snow, hail and any other is characterized by the accompanying complexity in their modelling. This is despite the need for quality and well-validated precipitation products, especially in the support of agriculture. There is growing literature and use of satellite-based precipitation data for drought monitoring (Funk *et al.*, 2014; Maidment *et al.*, 2014; Toté et al., 2015).

Satellite driven precipitation products are useful for drought and flood early warning systems and are meant to overcome the problem of limited distribution of rain-gauge observations and the tendency to have missing data from physical stations. The incompleteness of precipitation data may be due to damaged measuring instruments, changes to instrumentation with time, changes in data collectors and or change of measuring sites as documented in Sattari, Rezazadeh-Joudi & Kusiak (2017). There is, therefore, need to validate and document the accuracies of the precipitation products especially if physical ran gauge station data is to be used.

Toté et al. (2015) document the algorithmic approaches to the derivation of satellite-based precipitation products. These algorithms that derive precipitation from satellites are either *Thermal Infrared (TIR)* or Passive *Microwave (PM)* based. The basic assumption in the modelling of precipitation from satellites is the linear relationship between rainfall and cloud cover duration (CCD). Passive Microwaves are capable of penetrating clouds and thus capture better instantaneous rains and are more accurate over short periods compared to TIR based products that offer better accuracy over longer monitoring periods. TIR based algorithms are, however, susceptible to False Positives (FP) as a result of cold clouds with no rains like cirrus clouds and False Negatives on warm clouds not normally associated with rains like stratiform clouds but that then yield rains.

There is an increasing approach in combining both thermal infrared and passive microwave in precipitation monitoring. Examples of TIR microwaves includes Meteosat-8 and Geostationary Operational Environmental Satellite (GOES) while those of Passive Microwave (PM) includes (Special Sensor Microwave Imager, SSM/I; Tropical Rainfall Measuring Mission, TRMM and Advanced Microwave Sounding Unit, AMSU).

The current state of the art in the modelling of precipitation data from satellites involves the use of ground-based information to validate the modelled data. The blending with rain gauge data is meant to improve on accuracy and is documented for example in Sheffield (2014). Sheffield (2014), for example, use of rainfall data as achieving a 20% reduction of errors in P-values with 2 rain gauges per 1-degree box that is approximately a 111km box. A drastic reduction to 5% is recorded to be achieved with the use of 5 rain-gauges clearly indicating the benefit of use of more rain gauges in the calibration of satellite-derived precipitation datasets.

Modelling based satellite-derived precipitation data that use the validation approach include those from TAMSAT-Tropical Applications of Meteorology using SATellite (TAMSAT) in (Maidment *et al.*, 2014) and The Climate Hazards Group Infrared Precipitation with Stations, CHIRPS (Funk *et al*., 2014). There are current efforts, including the Global Precipitation Mission (GPM) Microwave Imager (MWI) to monitor both precipitation intensity and the 3D structure of rainfall particles through its Dual-frequency Precipitation Radar (DPR).

The common indices used for drought monitoring that are derived from Precipitation data sets includes, but are not limited to *Rainfall Estimates (RFE)* that is the absolute approximations of precipitation (Tarnavsky et al., 2014), *Standardized Precipitation Index (SPI)* (WMO, 2012) that is as standardised difference calculate from the general formulation in Equation (7) and *Rainfall Condition Index (RCI)* (Du et al., 2013) that is a relative range difference calculated following on Equation (8).

- ***Hydrological & Agricultural Indicators***

Hydrological and Agricultural drought indicators have a fine line in literature with quite a lot of overlaps. The various indicators in this category that are derivable from remotely sensed data include Land Surface Temperature (LST) (Wan, Hook & Hulley, 2015), Evapotranspiration (EVT) and Potential Evapotranspiration (PET) as documented in Running, Mu & Zhao (2017), Standard Precipitation and Evapotranspiration Index (SPEI) (Beguería et al., 2014), Soil Moisture (SM), Stream Flow Index (SFI). It is instructive to note that some of these like SM are modelled variables just like satellite-derived precipitation data. While Normalized Difference Vegetation Index (NDVI) is a direct measurement of vegetation conditions, the other hydrological variables based on groundwater and streamflow are measurements that are at times not directly provided through remote sensing approaches. In fact, some regions that experience perpetual aridity do not have the benefit of rivers/ permanent streams on which most hydrological drought indicators are based.

Remotely sensed Vegetation Indices (VIs) have been widely used in the monitoring of greenness of vegetation and indirectly for drought monitoring and even crop monitoring (Mainardi, 2011; McVicar & Jupp, 1998; Peters et al., 2002; Rembold et al., 2013; Rojas, Vrieling & Rembold, 2011; Unganai & Kogan, 1998). The most common of these indices used are those that transform spectral bands signals of sensor instruments to corresponding vegetation conditions. NDVI is the most common of the vegetation indices used for the above and also in monitoring changes in phenology, changes in land cover and land use and effects of global warming.

The NDVI measures/quantifies the relative abundance and activity of green vegetation and is correlated with chlorophyll. Most of the applications of NDVI above are based on the extraction of trends from NDVI time series data. The derivation of the NDVI is from the reflectances of two bands of a sensor and is calculated from Equation 11.

$$NDVI = \frac{(NIR-RED)}{(NIR+RED)} \qquad (11)$$

NIR and RED are spectral reflectance measurements in the red and near infra-red bands respectively. Healthy vegetation has higher NIR reflectance and low reflectance on the RED band since the same is absorbed for photosynthetic activity. High NDVI, therefore, implies green vegetation. NDVI 0.6-0.8 implies highly dense vegetation comparable to temperate and tropical rainforests; 0.2-0.3 implies moderate density (shrubs and grassland) while 0.1 and below signifies barren areas, rock, sand and snow (Quiring & Ganesh, 2010).

The NDVI has been utilized for many years to measure and monitor plant growth and vigour, vegetation cover and biomass production from multispectral satellite data. NDVI has been used in the monitoring of plants during the growing season since it indicates expected greenness for specified time periods in history. Healthy vegetation denotes favourable climatic and environmental conditions, while poor vegetation condition is indicative of droughts and diminished productivity. At the same time, the interaction between precipitation and vegetation is recognized and has been modelled. There however remains the need to model the lag effect of NDVI on meteorological drought (Zambrano et al., 2016; Zhang et al., 2013). There exist multiple sensors, satellites and data sources from which space monitored vegetation data can be sourced. Some of these sources for satellite driven vegetation indices are as discussed below with a summary of their characteristics as shown in Table 3.

Table 3: Sources and attributes of NDVI Products (Phenology, 2011)

Sensor	Satellite	Orbit Frequency	Years of Data	Spatial Resolution	Processed Time Step	Latency
AVHRR	NOAA series	Daily	1989-present	1 km	1-week, 2-week	~24 hours
AVHRR	NOAA series	Daily	1982-2006	8 km	3x monthly	N/A
Thematic Mapper	Landsat 4-5	16 days	1982-2011	30 m	By scene	N/A
Enhanced Thematic Mapper +	Landsat 7	16 days	1999-present	30 m	By scene	~1-3 days
VGT	SPOT-VGT	1-2 days	1999-present	1.15 km	10-day	~3 months
MODIS	Terra	1-2 days	2000-present	250/500m / 1km	8-day, 16-day	~7-30 days
MODIS	Aqua	1-2 days	2002-present	250/500m / 1 km	8-day, 16-day	~7-30 days

Despite the popularity of the NDVI, there are documented limitations of the NDVI index and its usage. The three main limitations include: (1) saturation of NDVI values, (2) difficulty in interpretation and (3) susceptibility to atmospheric interferences, especially of cloud cover.

The first limitation of NDVI is the saturation problem of the values as is articulated in Liu, Qin and Zhan (2012) and in Huete et al. (2002). The saturation problem arises because the NDVI is a ratio of the values of the red and infrared bands that are not linear in their relationship. The problem is a dynamic range expansion problem characterized by the decrease in the intensity of a spectral line especially in light of dense vegetation. In modelling, this is handled by transformation to ensure the indicator is linearized. There are popular alternatives to the NDVI that handle the problem of saturation. Such includes Fraction of absorbed photosynthetically active radiation, fAPAR (Meroni et. al., 2013).

Second, to the problem of saturation, is the inherent problem of difficulty in the interpretation of the NDVI values due to the range of the values from -1 to +1. An absolute NDVI value might not be a good indication of the exact occurrence of any deficiencies of vegetation cover. There is always then need to mitigated on this through the calculation of relative range indicators or ratios from the absolute NDVI values. Therefore, there exists an increasing use of NDVI values in the calculation of indices for drought monitoring. Such indices that are strictly NDVI derived includes Vegetation Condition Index (VCI) and Standardized Vegetation Index (ZVI). These are modelled as shown in Equations 12 and 13 (Klisch & Atzberger, 2016). The VCI (Kogan, 1990) reflects both spatial and temporal vegetation variability and also identifies the impact of weather on vegetation. The VCI is therefore appropriate for the monitoring of agricultural drought since it shows the variability of vegetation intensity for similar times in history (Zambrano et al., 2016).

$$\boldsymbol{VCI} = 100 * \frac{(NDVI - NDVImin)}{NDVImax - NDVImin} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (12)$$

$$\boldsymbol{ZVI} = \frac{(NDVI - NDVImean)}{std(NDVI)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots.. (13)$$

The VCI and ZVI (Equation 12 & 13) are calculated similar periods hence for the same time steps in history. The use of the time and space elements in the calculation of the VCI ensure that we compare only the comparables. Therefore, for monthly monitoring, every month's values will be computed based on that given month's min and max for VCI just as the standard deviation and mean are also time-step dependent for the calculation of ZVI.

Finally, and in addition to the problems of saturation and difficulty in interpretation, is the third limitation of the effect of noise on the NDVI due to both cloud cover and atmospheric interferences on the sensors that leads to the wrong quantification of vegetation greenness. The problem of cloud cover is for example documented in Park (2013). Image processing must thus take cognizance to eliminate the effects of clouds and shadows attributable to topography and saturation values in the numbers generated by the geometry of the satellite observation or as implied by interference by the presence of water on leaves.

2.7.3 Issues in the Use of Remote Sensing Data for Drought Monitoring

The issues in the use of Remote Sensing data can be broadly grouped into three categories: The problem of multiplicity of data sources, the availability of vast of volumes of data and the quality of data, especially as applied to drought monitoring. These issues are discussed in the context of predictive systems for drought monitoring.

2.7.3.1 Problems and opportunities in the multiplicity of remote sensing data sources

Efforts at drought monitoring exist in a context that is characterized by three main points of general convergence that includes:

- the agreement on the existence of different types of drought- meteorological, agricultural, hydrological and socio-economic droughts (UNOOSA, 2015) that was illustrated in Figure 2.5 above.
- the non-existence of an agreed single one fits all definition of drought with the monitoring systems, therefore, aiming to monitor droughts in a context where multiple definitions exist (Lloyd-Hughes, 2014).
- the existence of multiple sensors, satellites and data providers of the same drought monitoring data and information. These different data and information come in different formats and with different spatial and temporal, repeat frequencies and data usage policies.

The above realities led to the existence of a multiplicity of indicators and indexes meant to monitor different types and phases of drought. The differences in the definition of drought ensure that there can never exist a one size fits all specification of drought monitoring requirements. Use of multi-sensor data has been documented to offer improved accuracies and result in better inferences as compared to when a single sensor is used (Dalla Mura et al., 2015).

Even when the drought monitoring data sources are agreed on, there still exist subtle differences in the indicators and or indices that are then extracted from the sources for drought monitoring. The multiple sources can be used together in different approaches: data fusion (Dalla Mura et al., 2015; Khaleghi et al., 2013) and multi indicator applications both through indicators and indices. Generally, indicators and indices have no major differences except for the possibility that indices could have more than a single indicator in its definition.

The problem of multiple sources of data and multiple indices is, therefore, itself an opportunity for decision making as pertains choice of the following:

- **i.** The choice of what source(s) of data are appropriate for the problem at hand. This appropriateness is based on time-scales, re-visit frequencies and spatial coverage of the data
- **ii.** The desirable characteristics of the drought to be subject of monitoring including severity, duration and spatial extent.
- **iii.** Availability of historical records and ease of access or computations of the same.
- **iv.** The ability of the data to support near real time computation and thus incorporation into early warning systems.

Approaches to address the Multi-source data problem in drought monitoring

The multi-source data problem raises the need to understand the commonalities between the multiple data sources, multiple indicators and indices. This is so, due to the existence of an array of indices and data sources that are in the vast competitive, complementary and or independent. Approaches to handling multiple datasets in the case of drought monitoring include the use of a single indicator/ index or the use of multiple indicators/indexes.

- ***Use a single indicator /index*** is the premise of most drought monitoring systems. The basis of this approach is the use and reliance upon a single drought indicator or index. The use of a single indicator can be viewed as a naïve approach while that of a single index could be viewed as a data reduction and simplification approach for ease of communication. An example is a recommendation by the World Meteorological Organization (WMO) on the use of the Standardized Precipitation Index (SPI) as the unifier index for drought monitoring (Hayes et al., 2011). Klisch & Atzberger (2016) also documented the use of the Vegetation Condition Index (VCI) in a single indicator drought monitoring model. This approach, therefore, remains easy to interpret and communicate but would not make for an effective decision support system as in the most, it covers only one type of drought.

- ***The use of multiple indicator/indices*** for drought monitoring is becoming increasingly applicable in drought monitoring. In this approach, multiple indicators/ indices are used to monitor drought either of different types or of drought in its entirety. The approach involves either the calculation of a hybrid/super index that combines multiple indices or the use of multiple different indices in a multivariable setup. Such approaches are documented in different studies:
    - Enenkel et al. (2016) use the Enhanced Combined Drought Index (ECDI) that integrates four input datasets: rainfall, moisture, land surface temperature, and vegetation status. The datasets are weighted for each pixel with an automated redistribution of weights for cases when missing data is encountered in any of the component datasets. This study proposes the combination of this dataset with socio-economic data sourced using smartphones from the communities.
    - Hao & AghaKouchak (2013) used the Multivariate Standardized Drought Index (MSDI) that probabilistically integrates the Standardized Precipitation Index (SPI) and the Standardized Soil Moisture Index (SSI) for drought characterization. The key reason advanced by the study for use of multiple indexes is the insufficiency of a single index to reliably assess drought risk and serve for decision making. The approach can, therefore, be referred to as multivariate, multi-index drought-modelling.
    - Vicente-Serrano et al. (2012) that compares the Standardized Precipitation-Evapotranspiration Index (SPEI) and the Standardized Precipitation Index

(SPI) over other Palmer's drought indicators. The study documents the superiority of the performance of SPEI & SPI over the Palmers indices.

- Sun, Mitchel & Davidson (2012) proposed and used a Multi-Index Drought (MID) model that combines various indicators for agricultural drought in the assessment of wheat crop yields. The study reported the superiority of the MID models over single indices/ indicator models.
- Zhu et al. (2016) documents the use of both the Standardized Precipitation-Evapotranspiration Index (SPEI) and the Standardized Precipitation Index (SPI) calculated over a period of 1 to 12 months to detect hydrological droughts. Use of multiple time-scales realized better probability of detection of hydrological droughts making for a good alternative when streamflow data is not available.
- Touma et al. (2015) use a multi-model and multi-index approach to the evaluation of drought characteristics. Data from 15 climate models from and multiple indices are used to assess the likelihood of changes in the spatial extent, duration and number of occurrences of future droughts. The four drought indices: the Standardized Precipitation Index (SPI), the Standardized Runoff Index (SRI), the Standardized Precipitation–Evapotranspiration Index (SPEI) and the Supply–Demand Drought Index (SDDI) are used.

The three main issues around the use of multiple indices include the assurance of continued availability of the multiple datasets from the multiple sources, the ease of interpretation of the resultant index and the handling of the computational complexity that in most cases are part of their derivations.

In the context of drought monitoring, the simplest application of remote sensing data has either of or both of precipitation and vegetation-based indexes. Even in these simple applications of the remote sensing is afflicted with the diversity that comes with the precipitation and NDVI products. Meroni et al. (2013) have discussed these factors as spatial and temporal resolution and the availability and quality of data together with the intended application and application areas into which the data will be deployed.

As an illustration, a consideration of vegetation datasets documents several satellites that have provided large scale monitoring for vegetation. Several NDVI datasets with global coverage are available borne out of diverse sensors and algorithms. Possible sensor sources of NDVI data include AVHRR NOAA, MODIS NASA, SPOT VEGETATION (VGT) and SeaWiFS (Meroni et al., 2013; Scheftic et al., 2014; Wenxia et al., 2014). There have been undertakings to construct NDVI time series that extend their coverage to the early 1980s.

In the cases where multiple data sources exist, comparison for purpose and objectivity in choice of datasets becomes a key undertaking. The choice is therefore between products with competing characteristics. The use of multiple indices or selection of single indices from several possible multiple sources is mainly driven by the use of data comparison techniques that are then geared towards ensuring objectivity in the choice of datasets for drought monitoring. The available datasets should thus be evaluated for similarity and divergence like is the case in Albarakat & Lakshmi (2019) and Martínez-Beltrán et al. (2009). Such identified similarities or divergences can inform the best course of the use of the multiple datasets which could be either the selection of a single one or use the multiple sets of data.

The methods documented for the investigation of similarities and differences between similar remote sensing datasets includes a comparison of distributions, comparison of correlations and comparison of agreements. These are summarized as follows: -

Comparison of distribution

The use of distribution functions to describe similarities or differences between data sets is a widely employed method. The four characteristics of data that determine a choice of distribution are possible data values based on whether the data values are discrete or continuous; symmetry and direction of the symmetry that indicates the presence of both positive and negative outliers; the existence of upper and lower limits on the data for example if between 0 and 100 and the likelihood of observing extreme values based on the frequency of the extreme values. The comparison of NDVI datasets from multiple sources is for example documented in Yin et al. (2012) for MODIS, AVHRR and SPOT-VGT.

The approaches to the statistics of distribution can be classified based on whether distribution parameters are known beforehand on the data. This gives either parametric or non-parametric methods. ***Parametric Methods*** are those that make inference based on the assumption of parameters to a distribution function. Such is the basis of use of the standard distributions including Binomial, Poisson, Geometric and Discrete uniform distributions. Parametric methods, therefore, rely on the tremendous reduction of original problems to a few parameters. This reduction is achieved by making many and mostly over-restrictive assumptions. They are convenient when correct, efficient and easy to interpret. ***Non-parametric methods,*** as opposed to parametric methods, non-parametric methods make as few assumptions as possible. A distribution form is thus not defined over a function F(x) as long as it is cumulative distribution function. This approach, therefore, leads to the approximation of a function as opposed to a parameter. One non-parametric approach that is commonly used to visually compare data distributions is the Empirical Cumulative Distribution Function (ECDF). The ECDF is defined as shown in Equation 14: -

$$F(x) = \frac{1}{n}\sum_{1}^{n} I(x_i x) \ where\ I\ is\ the\ indicator\ function \ldots\ldots\ldots\ldots\ldots \quad (14)$$

F is the CDF function and is noted to put a mass of 1/*n* at each data point x_i

Comparison of correlation

Correlations are used to describe the existence of a relationship between variables, with the concept extendable to cover datasets. The investigation of correlation is quite popular in the investigation of data archives for both agreement and differences. The correlational analysis is for example used in Yin et al. (2012) to analyze differences and agreements among MODIS, AVHRR and SPOT-VGT datasets. The use of NDVI trends in Yin et al. (2012) in our opinion is essentially the same approach to the investigation of correlations despite the study treating them as different. The correlational analysis approach is also used in Song, Ma & Veroustraete (2010) to validate the linear relationship between the two types of NDVI products from SPOT-VGT and AVHRR sensors. The most common methods of specifying correlation are Scatter Plot and Correlation Coefficient as discussed here next:

Scatter plots that have two data sets plotted on a graph of paper along the same x and y axes. A visual inspection can be used to decide whether the correlation, r is perfectly positive and therefore r=+1 and all points lie on a straight line, correlation is perfectly negative when r=-1 and any dispersions based on their directions result in either high/low degree positive/negative correlations. The base case, r=0 implies either a broad spread over a broad area with a downward trend or absence of correlation. The method is widely documented to be non-mathematical, naïve and un-reliable without the ability to measure the degree of correlation.

Coefficients of Correlation is defined by Equation 15. Denoted as r, the linear correlation coefficient measures the strength and direction of a derived linear relationship between two variables. The interpretation is much like the Scatter plot, but with a mathematical quantification and direction specification of correlation.

With a value between -1 and 1, the interpretation is that a value approaching -1 implies a strong negative correlation, 0 no correlation and those approaching +1 have a strong positive correlation. Heuristically, a 0.7 cut off signals a strong correlation while below 0.5 implies a weak correlation. For the formulation of magnitude, the determinant of correlation- r^2 or R^2, is always used. R^2 implies the proportion of variations of y explained by the linear relationship between x and y. The limitation in correlations is that we only measure linear co-variation and not actual difference.

$$r = \frac{n\sum(xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \quad (15)$$

Comparison of agreement

Agreement methods measure agreement between two variables and by extension datasets. The bounds of the agreement coefficient (AC) is defined between 0 and 1 for no agreement to perfect agreement for this measure thereby making it easy to interpret. The AC as documented in Ji & Gallo (2006) is noted to be non-dimensional, bounded, symmetric and distinguishable between symmetric and un-symmetric differences. The AC is defined in Equation 16.

$$AC = 1 - \frac{SSD}{SPOD} \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{(16)}$$

Where, SSD is the sum of square difference (SSD) and SPOD is the sum of potential difference (Meroni et al., 2013). SPOD is defined by Equation 17.

$$SPOD = (\quad |X_i - \bar{X}| + |\bar{X} - \bar{Y})(\quad |Y_i - \bar{Y}| + |\bar{X} - \bar{Y}) \text{ .... (17)}$$

2.7.3.2 Problem of availability of vast volumes of remote sensing data

There is an explosion in the availability of remote sensing data at volumes and frequencies that qualify it for big data. Big data was previously characterized by three attributes: *volume* in the order of exabytes of data, *velocity* based on a very high frequency of incidence, *variety* in different formats- both structured and unstructured as initially documented in Russom (2011). Additional characteristics now documented in big data include veracity and value (Anuradha, 2015). While *veracity* raises quality issues on the data by positing questions on completeness, cleanliness and accuracy of the data, *value* is perhaps the most important as it poses the question of business value derivable from the data. The specific case of climate data, that is the basis of this study, has complexity in presentation and storage as an attendant characteristic.

The availability of vast volumes of data covering long periods and at regular frequencies and with expected availability in the future makes an opportunity rather than a limitation. With the enhancement of big data analysis techniques, the discovery of nuggets of importance from these data makes for a perfect convergence of data and tools. The data repositories form time series data that can be subjected to time series analysis and even prediction using machine learning techniques. The two common models for time series analysis are the Brockwell and Davis generic model and the Fourier series model.

- **Brockwell and Davis Generic model**

One generic model that remains popular for time series analysis is the model described in Brockwell & Davis (2006) that is summarized in Figure 2.6. Although the steps were presented as sequential, our understanding of the model is that it has an inherent back and forth mechanism between the subsequent stages of the process.

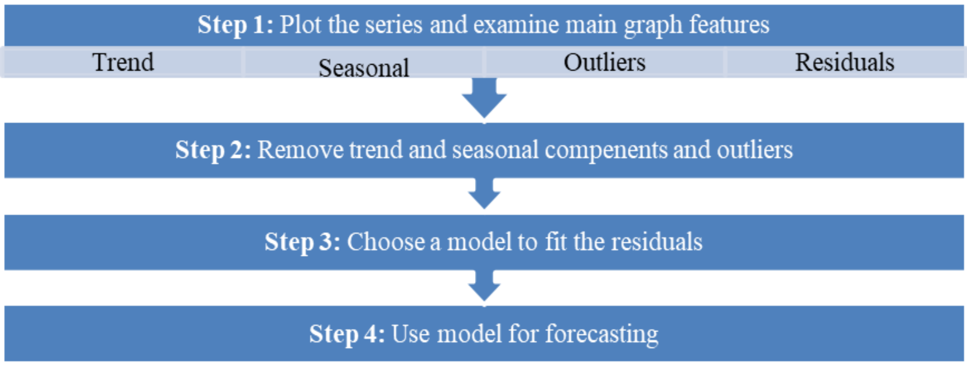


Figure 2.6: The Brockwell and Davis Model for time series modelling (Brockwell & Davis, 2006)

The Brockwell & Davis model aims to realize the stationary components of the time series that are referred to as residuals even in cases in which it is necessary to do transformations. The choice of model to fit residuals uses many sample statistics like the autocorrelation function.

Data for drought monitoring, both Remote Sensing and Socio-Economic data, is generally collected in regular time intervals that could be any of daily, weekly, dekadal, monthly, quarterly, annually etc. The data, therefore, has a temporal dimension on top of the spatial dimension that is majorly inherent in the data. The regularity of the data collection makes this data to be interpretable as time series data. Brockwell & Davis (2006) define a time series to be a set of observations x_t, each one being recorded at a specific time t. This definition is extended to define discrete time series to be that in which the time intervals, t, are discrete. Drought monitoring data are thus, generally, discrete time series datasets.

- **Fourier series model**

An alternative model to the Brockwell and Davis model is the resolution and expression of the time series as a component of its Fourier components. The modelling expresses the time series as a set of waves with different frequencies.

Temporal domain data can be transformed to an equivalent frequency domain using by the use of Fourier analysis (Moody & Johnson, 2001). For discrete data like that of the remote sensing time series data, a discrete Fourier transform (DFT) is used as long as there exists regularity of spacing of the data points in the temporal domain.

The use of the above models in the analysis of multiple datasets is, in essence, a data comparison problem that aims to realize the differences and similarities between multiple time series data. The methods of data comparison have the objective of realizing insights on the distribution, correlation or agreement between the data items (time series) of comparison. Below we describe some of the specific approaches to the comparison of the different datasets and time series that fall within the generic model of Brockwell & Davis (2006).

The processing and analysis of time series data are primarily geared towards generating models that are meant to achieve some specific objectives. These objectives include those of ***decomposition of the time*** series though the extraction and separation of the time series into trend, seasonal and random components as documented in Brandt et al. (2014). The decomposition is based on a locally weighted regression smoother filter. The season term is generally dropped off from the long-term analysis.

Further to the decomposition of the time series data ***is noise filtering*** that involves the removal of any un-intended but captured data that lead to effects and biases on the captured values. The objective of the ***prediction of future values*** based on the existing values is meant to offer support for early and real time or near real time monitoring.

Accompanying the decomposition of datasets are the twin concepts of hypotheses testing and simulation for the generation of new insights. Testing some given hypotheses such as increasing frequency of droughts over the years for some given spatial coverage is meant to provide evidence for areas that need closer attention while the simulation of the data in novel ways is meant to lead to the generation of new insights.

Essentially, the problem of multiple time series data comparison for similarities and differences has been widely studied. In its simplicity, the comparison of multiple time series data can be defined as that of seeking a constant model with a goodness of fit that is capable of accounting to the difference and/or similarities between different time-series data (Jin, 2011). The wider goal is to account for the differences and similarities and extend these to multiple time series that could, in essence, be of different lengths.

2.7.3.3 Problem of quality of data in remote sensing

Data quality, as opposed to say quality as understood in manufacturing, is defined in terms of intangible characteristics as opposed to physical properties. The issue of data quality is increasingly becoming crucial in remote sensing for a trio of reasons: (1) is the fact that many non-government entities are getting into space and deploying satellites; (2) is the increasing use of remote sensing data for decision making even in critical applications and; (3) is the continued reliance on digital technologies and thus secondary sources for the data as advanced by Batini et al. (2017).

The concept of data quality is best described in-line with the key attributes of resolution (Lefsky & Cohen, 2003). Types of resolution include: the *spatial resolution* that defines pixel sizes, the *radiometric resolution* that defines the different number of intensity values in an image, the *spectral resolution* that defines the number of channels recorded and *temporal resolution* that defines the frequency of the data capture and thus a major basis for monitoring systems.

Figure 2.7 illustrates the data quality definitions for remote sensing data based on the above resolutions- spatial, radiometric, spectral and temporal using a 3x4 image with a total of 12 pixels for each band.

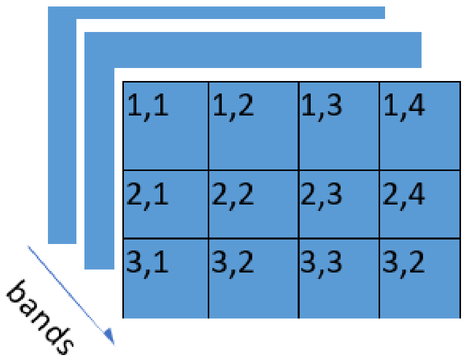


Figure 2.7: A sample illustration of a 3-band 3x4 remote sensing image.

Remote sensing images are raster images provided as matrixes with a given number of rows and columns. While *geometric or spatial precision* is a definition of the homogeneity of the pixels for all the bands, *radiometric precision* concerns the representativeness of the digital values stored for each pixel as a measure of reflectivity. The accuracy in the separation of the bands defines the *spectral precision* while temporal *precision* would imply the metadata on date and time of capture remain accurate. Finally, as *positional precision* defines the image and its relation to a coordinate system, attribute accuracy is indicated by the goodness of measure of thematic interest.

Assuming the correct set up of both satellites and sensors, the most common causes of data quality issues in remote sensing images is the loss of radiation as a result of interaction with the atmosphere. One such cause of loss of radiation is the effect of clouds in the obstruction of radiations from interacting with objects of interest. Handling effects of atmospheric interference is key in remote sensing data since it is the biggest cause of radiometric inaccuracy of remote sensing images. There have been documented cases of some areas of interest being covered by clouds up to 67% of the time within the year (Wang et al., 1999).

Data smoothing and filtering as well elaborated in Klisch & Atzberger (2016) are especially important in the processing of remotely sensed data particularly for vegetation that is in the most cases affected by cloud cover and other atmospheric interferences. In this aspect, the processing of Remote Sensing data could benefit from data mining models that have been advanced for their ability to fill in the gaps in data by extrapolating and estimating necessary parameters.

2.7.4 Socio-Economic Data for drought monitoring

Despite the documented socio-economic effects of drought, not much documentation exists in their modelling within the practice of drought monitoring. This is perhaps because of their tendency to be affected late in the drought cycle or because of their affinity to have their trends affected by interventions in the form of drought response and mitigation that serve to minimize impacts of droughts on communities. A good review of interventions in the pastoral livestock sector and their possible socio-economic benefits is provided in Morton et al. (2005).

The effects of drought are generally considered in the three categories: social, economic and environmental effects. Jenkins (2012) and Garrido (2014) document the Hochrainer model (Figure 2.8) that describes the categorization of the impacts of drought based on the three widely used categories of society, economy and environment. An alternative to this common classification views drought in terms of the sectors affected, effects on supply and demand, whether impacts are tangible or intangible and the effects on the environment.

Direct, indirect or secondary micro-impacts of products are realized based on the sectors that are affected by the drought. Direct effects are those that mainly impact the productive sectors of crop agriculture, livestock production, fish production etc. while indirect impacts are generally results of the direct impacts like effects on the agricultural food sector. While n*on-market impacts like welfare reductions* are due to effects on demand and supply markets (Garrido, 2014) that then lead to effects on access and subsequently impact the welfare of communities under exposure; *Tangible and intangible impact*s is an alternative classification offered by Massarutto et al.

(2013) that in addition to direct and indirect impact, also considers the impacts on lifestyle, health and social tension. On the other hand, *environmental impacts* include those on water systems (both surface and groundwater) and effects of wetlands at the occurrence of extreme droughts.

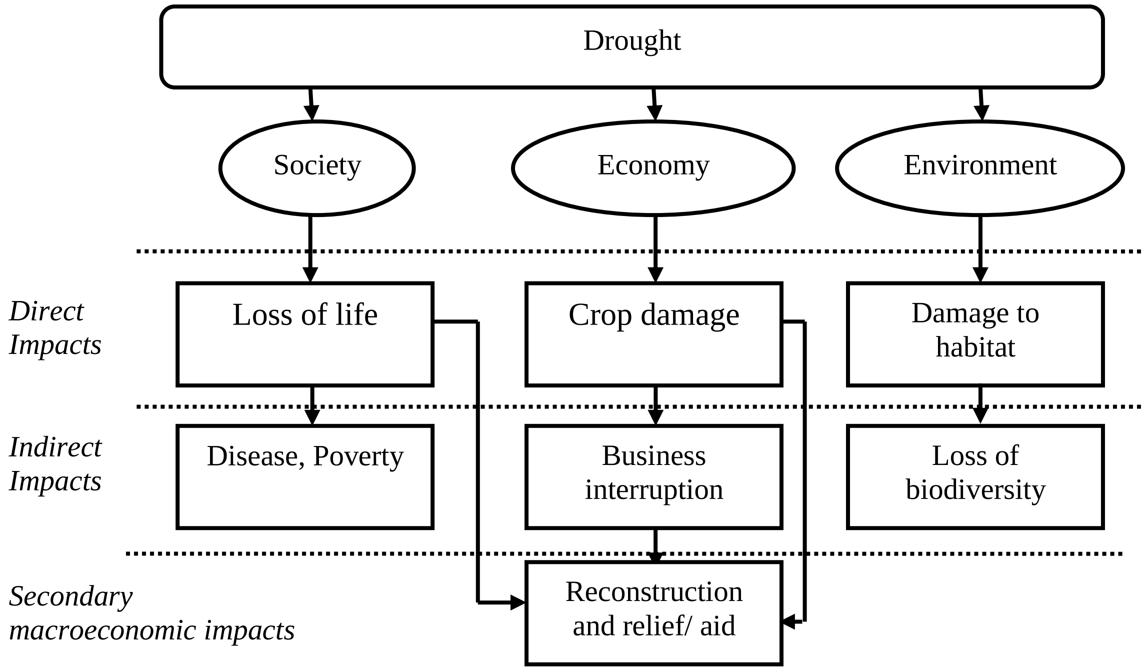


Figure 2.8: The Hochrainer illustration of the impact of drought showing the social, economic and environmental impacts of drought (Jenkins, 2012)

Given that droughts have both a spatial and temporal coverage, impacts of their occurrence on these extents can be monitored by proxy indicators that quantify the impacts of the droughts on the socio-economic conditions of the areas of coverage. This approach, as documented in Massarutto et al. (2013), should involve the twin requirements of the *quantification* of the socio-economic and environmental effects of drought events and the *modelling* of the relationship between the socio-economic impacts and selected variables that monitor and are associated with drought.

The limitations to the use of Socio-Economic data in drought monitoring are majorly due to three broad issues around the identification, documentation and quantification of droughts and the losses attributable to their impacts. While environmental and even socio-economic impacts of drought are generally difficult to identify and quantify in monetary terms (Ding, Hayes & Widhalm, 2011), intangible impacts of drought are even more difficult to identify. The quantification of non-tangible impacts of drought is therefore guaranteed to be a non- trivial task.

2.7.5 Socio-Economic Indicators for drought monitoring

Just as the case for remote sensing data, to be used in drought monitoring and drought prediction, socio-economic data has to be transformed into indicators that are capable of correlating to drought. Such transformations should aim to make the datasets sensitive to changes in drought severity. The classification of impacts of drought on socio-economic terms follows the model provided in Figure 2.9 that categorizes the impacts into those on production, access and welfare.

The categories of socio-economic data that can be collected for drought impacts include the impacts on crop and livestock agriculture, access indicators like market prices and welfare indicators like nutrition-based indicators.

Figure 2.9: Sequence of effects of drought on livelihoods as drought progresses.

Changes in drought severity or enhanced exposure periods lead to an impact on both crop and livestock production. Access to produce is then affected as a result of high demand in the context of reduced supply for particular goods and services. The welfare of the communities is the last aspect to be affected, especially with malnutrition rates getting escalated to beyond normal levels.

The common production indictors for the production effects of drought include crop yields and milk production. The market access indicators are for example documented in FEWSNET (2009) and in Hill & Fuje (2017) and include those that measure changes in staple prices e.g. of maize, prices of livestock and terms of exchange between staples and popular production holdings. The effect of droughts on food prices based on survey data is further documented for the case of a Kenyan county in Mohajan (2014) and for hay and feed prices in Schaub & Finger (2020). One study that stands out in investigating the relationship between vegetation conditions and maize prices is found in Shuaibu et al. (2016). The regression model in Shuaibu et al. (2016) concludes that the normalized difference vegetation index (NDVI) is a good index for modelling the change of maize prices and is hence useful or emergency planning.

The popular welfare indicator on malnutrition, the Middle-Upper Arm Circumference (MUAC) as documented in De Onis, Yip & Mei (1997) and in James et al. (1994) is best measured in the sub-set of the population below 5 years.

2.7.6 Methodologies of Handling Socio-Economic Loses from Drought Impacts

Drought occurrences, especially in extreme cases provide a lot of shocks to communities as modelled by the Hochrainer model in Figure 2.8. There exist varied ways of responding to these shocks including offering no protection to communities, building the resilience of communities, enhancing drought preparedness and drought response and drought insurance.

- ***The no protection of communities:*** Is an approach that is characterized by the lack of planning coupled with the unwillingness to cushion societies from losses. The communities are therefore left to bear the consequences of droughts. This approach is laden with massive losses of both life and livelihoods.
- ***Resilience building:*** Resilience building to extreme events remains quite complex (Tortajada et al., 2017). Basically, resilience-building involves improving the capacities of communities to handle drought shocks for increasingly longer times and at greater impacts. This is the current method of choice that is geared towards making communities self-reliant and therefore prone to suffering less drought effect on both lives and livelihoods.
- ***Contingency planning, drought preparedness and drought response***: Is an approach that is driven by planning, in advance, for droughts. The realized advance plans are referred to Contingency Plans and they have, in general, well-modelled scenarios. The contingency plans are coupled with drought preparedness that is similar to resilience building. In the event that droughts

occur, drought response involves taking actions that protect communities from the loss of lives and livelihoods.

- ***Drought Insurance:*** is an approach that is increasingly becoming popular as an option for handling socio-economic losses from droughts. In this case, premiums are agreed and paid for before-hand against possible drought episodes, especially for agricultural droughts. Insurance risks against droughts are considered a systematic risk and thus valued expensive. This approach typically involves the use of an objective monitoring system, that is typically driven by remote sensing, for the quantification of even occurrence and socio-economic data for determination of impacts and thus levels of pay-out. An operational drought insurance system that is based on the use of remote sensing data in the insurance of drought losses is, for example, documented in Mude et al. (2010) and in Chantarat et al. (2013). It is, therefore, a technical undertaking to develop such specialized insurance schemes against, not only drought but other natural disasters.

2.8 Formulation of the drought prediction problem

2.8.1 The generic prediction problem and drought

The drought prediction problem is a sub-set from the domain of prediction problems. An investigation of the drought prediction problem, therefore, follows from the definition of the prediction problem. The general formulation of the Machine Learning approach to the definition of the prediction problem is provided in Figure 2.10.

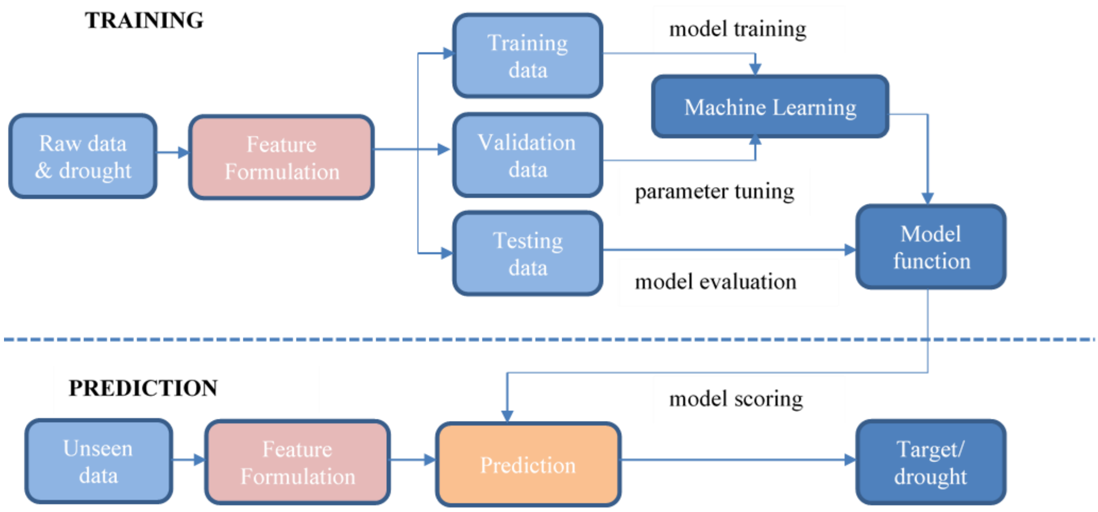


Figure 2.10: The machine learning approach to prediction as adapted from Mitchell (1997).

The model building process has at the core the use of historical datasets to formulate models and that are then used to make predictions. The generic prediction model has several steps that are geared towards realizing models capable of making predictions. The main tasks are feature selection, model training and model parameter approximations through model validation and the evaluation of the performance of the models. The model function realized is finally used in the approximation of the target i.e. drought in the case of this study.

From the generic machine learning (ML) model provided and in the context of drought monitoring, two tasks are non-trivial. First is the definition of drought as the target variable. The question is best answered by any of the Equations 3-10 (Table 2) that outlined the common difference indicators used in drought monitoring. The second task is the choice of the ML algorithm to be used in learning the model that best approximates drought. The methods are discussed in sections 2.8.3-2.8.5 including the possible methods for both parameters tuning and model evaluation.

Important to observe from the generic model is the attendant limitation that any drought monitoring process would have to overcome. This limitation is on the application of the model for scoring. In the ML problem, we predict future conditions using past recorded conditions. In a strict sense, it is not like the generic problem where the features of the instance are provided. In fact, the future remains strictly defined by the past. The concepts used in the prediction of future conditions are discussed in section 2.8.2. The closely related concepts in prediction are Machine Learning (ML), Knowledge Discovery from Databases (KDD), Data Mining and Artificial Intelligence (AI). Section 2.8.3 zeroes in on the ML methods for approximating the model function.

2.8.2 Machine Learning (ML), Data Mining, Artificial Intelligence (AI) and Knowledge Discovery from Databases (KDD).

Closely related to Machine Learning (ML) are the concepts of Data Mining, Knowledge Discovery from Databases (KDD) and Artificial Intelligence (AI). They, however, are slightly different concepts. The confusion is majorly on the twin pairs of Data Mining and KDD; and AI versus ML. A highlight of their differences is presented between the pairs most closely related as discussed in sections 2.8.2.1 to 2.8.2.3.

2.8.2.1 Data Mining & Knowledge Discovery from Databases (KDD)

Data mining is defined as the extraction of useful models of data either in the form of summarization of the data or identification of extreme features of data (Han, Kamber & Mining, 2006; Hand, Mannila & Smyth, 2001). This definition of data mining is closely related to that of Knowledge Discovery from Databases (KDD) that is defined with the underlying concept as the exploration and analysis of large quantities of data in order to discover valid, novel, potentially useful and ultimately understandable patterns (Fayyad, Piatetsky-Shapiro & Smyth, 1996; Goebel & Grunewald, 1999; Han, Kamber & Mining, 2006). Although close in definition, it is the case that KDD is a wider area of which data mining is a sub-process. KDD includes knowledge extraction and representation at the tail end of its process. It is our understating that data mining is best described as part of a process that aims to find patterns from data while KDD is the super-process that also grapples with how to represent knowledge and reasoning on the knowledge.

2.8.2.2 Machine Learning & Artificial Intelligence (ML versus AI)

Artificial Intelligence (AI) aims to create machines that are capable of mimicking both the human mind and behaviour and that also has learning capabilities at its core. Therefore, AI encompasses more than the concept of learning and includes the concepts of knowledge representation, reasoning and abstract thinking. AI is therefore different from machine learning that is more focused on the creation of software systems that can learn based on past experience.

2.8.2.3 Data Mining and Machine Learning

Data Mining is the process of identifying new patterns and insights from data. Data Mining thus involves the extraction of regularities from very large datasets/ databases as part of a business application process (Fürnkranz, Gamberger & Lavrač, 2012; Kohavi, 2001). The key undertaking in Data Mining is thus the extraction of interesting and thus non-trivial, implicit and previously unknown nuggets of potentially useful information from data resident in large databases.

Data Mining is thus driven by four main factors that include: the data availability factor, the need for interpretation factor, the need for prediction of the future and the availability of storage and related technologies.

The existence of vast volumes of data ***(***Anuradha, 2015) that is either structured or unstructured is a drive to data mining. There is an increase in the variety of automated data collection in diverse areas including remote sensing (Liu, 2015) that then lends itself to data mining that generally is considered to be data-hungry. The need to make sense of the above vast volumes of data especially that from operational systems and data warehouses is an increasing undertaking of businesses. This is further supported by the advancements in data storage and methodologies and tools for data analysis, summarization and visualization that have made it easier to use data for decision making. The need to attain the twin goals in data mining of gaining insights from data and using the same to predict the future based on the existing past data especially for comparative advantages is a key driver to data mining. The data collected from such is variously documented as the next opportunity for not only productivity but also competition and innovation (Manyika et al., 2011; Zikopoulos & Eaton, 2011).

Machine Learning (ML) is a close concept to Data Mining that is operationally defined by Mitchell (1997) in terms of the ability of a machine to learn from experience. A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P if its performance at Tasks in T as measured by P improves with experience E. The experience E is provided in terms of a training dataset over the task T. As variously defined (Bishop, 2006; Mitchell, 1997; Nilsson, 1996), the overriding concept in ML is the ability of the learning techniques used to not only predict an outcome given some input but also to improve at the task given more experience.

There exist subtle differences between Data Mining and Machine Learning. The differences are based on focus, data requirements, and goal orientation of both Data Mining and Machine Learning. First, whereas Data Mining focuses on the discovery of previously unknown properties of the data, Machine Learning (ML) focuses on known properties learnt from the previously existing data. Second, while Data Mining is mostly driven by the existence and use of large datasets, ML though a potential beneficiary from large datasets possesses algorithms that also lend themselves to handling small data sets. The final difference is premised on the fact the Data Mining has an overall goal of finding nuggets of information from the huge sets of data. This search is not particularly based on pre-set and guided rules and goals. Data Mining can, therefore, be terminated at the exploratory stages given it is non-specific in goal orientation on the data while ML can be viewed as goal-oriented in the search for specific outcomes.

Recent developments have seen the adoption of Data Mining across many industries for diverse applications. This widespread adoption of ML approaches has then raised the need for standardization of process of Data Mining, the result of which is the Data Mining Process.

2.8.2.4 The Data Mining Process

The data mining process is variously described (Azevedo, 2008; Chapman et al., 2000; Fayyad, Piatetsky-Shapiro & Smyth, 1996). Despite the differences in the description of the Data Mining process, the common steps are summarized in the Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology (Azevedo, 2008; Chapman et al., 2000).

The CRISP-DM process is widely accepted and aims to simplify the Data Mining process to a step of processes as modelled in Figure 2.11.

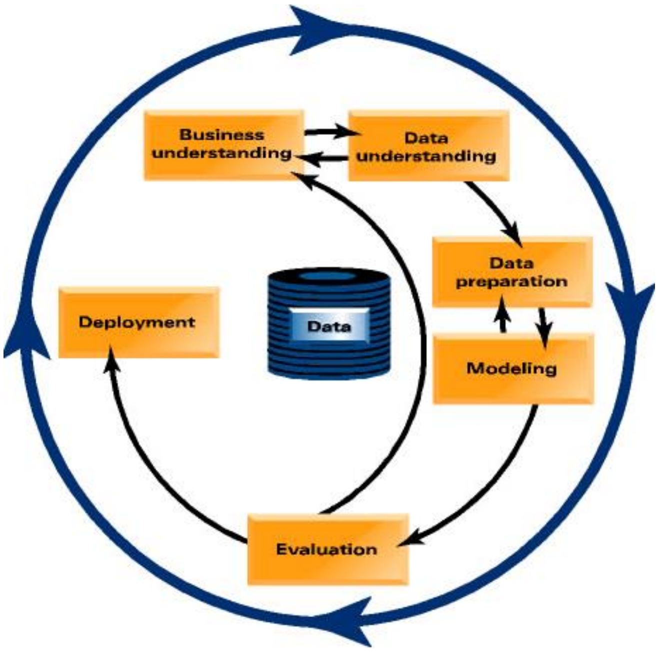


Figure 2.11: CRISP-DM Data Mining Process Model.
Data is at the core of the modelling process of the methodology that also provides for model deployment.

The data mining process (Figure 2.11) follows the steps outlined below: -

1. **Business Understanding** that encompasses working with business to define objectives from requirements and ends with a well-formulated data mining problem to address the objectives of the solutions sought.
2. **Data Understanding** to formulate a hypothesis after getting a preliminary understanding of the data and its associated problems.
3. **Data Preparation** in which the data is converted, through transformations and attribute selection, to a final form from which a model will be developed.
4. **Modelling** stage that applies selected modelling techniques and is mostly interactive with the data preparation phase.
5. **Evaluation** sub-set at which the model is assessed for performance and if its results meet earlier set objectives and apply to the business objective. It is the case that a choice has to be made from the use of multiple models.

6. **Deployment** is a non-trial task involving the actual use of the models to offer solutions to the earlier defined problem. Deployment takes various forms from simple communication of results to deployment of management tools to accompany the models.

An alternative to the CRISP-DM methodology is the SEMMA methodology. SEMMA is an acronym for steps in this model for data mining that follows on: Sample, Explore, Modify, Model, and Assess phases for a typical data mining process. Although variously defined as a methodology (Azevedo, 2008; Mariscal, Marbán & Fernández, 2010; Nadali, Kakhky & Nosratabadi, 2011), SEMMA could be considered as a logical organization of the data mining process. The Key steps of the methodology are in most cases iterative and thus should not be linearity assumed in the progression of the data mining process steps. The SEMMA methodology phases are as follows: -

1. **Sample& Explore: -** Whereas the Sample process is an optional process that involves retaining only the data that is considered useful, the explore process supports the discovery process on the data and can involve the use of both visual and statistical techniques.
2. **Modify and Model: -** The Modify phase involves creation, selection and transformation of variables and is succeeded by the Model phase. The modelling phase involves the use of modelling techniques to combine variables to predict useful outcomes.
3. **Asses: -** Typically, this involves the evaluation of the usefulness of findings from a model intending to estimate how well the model performs. The most common approach is the application of the model on both test and validation datasets that are different from the data on which the model is trained.

It is the case that the SEMMA methodology does not explicitly provide for the key step of data pre-processing. However, an alternative view could be that data pre-processing is implicitly provided for within the Sample phase of the methodology. The application of the SEMMA based methodologies will thus have to be modified to include an extensive phase for pre-processing of the scientific data. On the other hand, the CRISP-DM methodology is elaborate enough to provide for all the steps required to build models on climate data since data pre-processing is extensively provided for. Viewed closely, it is noted that, despite the differences in number and sequences of stages, the CRISP-DM and SEMMA methodologies, in essence, remain a description of the same process.

The discovery of hidden nuggets is a key undertaking of data mining. Mena (1999) documents data mining as involving the discovery of actionable and meaningful patterns, profiles and trends by sifting through data using pattern recognition technologies such as Neural Networks and other machine learning algorithms including genetic algorithms. Openshaw (1999) also asserts the appropriateness of data mining in the handling of vast volumes of data as is the case with sensor data in drought monitoring. The possibility of the use of data mining and machine learning techniques in data reduction and data visualization cannot be over-emphasized. This is made possible by the suitability of machine learning algorithms in the modelling step of data mining. The automation of the data reduction process is one benefit that drought monitoring can realize from the combination of Data Mining and machine learning.

2.8.3 The Machine Learning methods of drought prediction

Having formulated the drought prediction problem as a machine learning problem that uses past/ historical data to approximate a function that can be applied to predict drought, the question then shifts to the possible methods that can be used for the prediction.

Machine learning takes the form of any of the following popular and commonly used three learning paradigms of supervised learning, unsupervised learning and Reinforcement Learning.

- ***Supervised Learning*** has methods similar to those outlined in Bishop (2006) and Mitchell (1997) in which models are presented with training data in the form of labelled training examples. The machine in this instance is presented with observed data that have a specified outcome referred to as output labels to derive a function that it then uses to approximate the outcome of previously unseen and thus unlabeled instances. Supervised learning tasks are either classification or regression tasks based on the outcome to be approximated. If the outcome of the training to be approximated is discrete the task is a classification task, while regression tasks approximate continuous value outcomes. Supervised learning, being more structured, remains the most common approach to ML since it is possible to subject it to evaluation for performance.

- ***Un-supervised Leaning*** is when the training instances have no labels and the machine is to look for relationships from the dataset provided. We assert that Unsupervised Machine Learning is in the most similar to descriptive data mining through which useful insight is gained. Unsupervised learning methods rarely lend themselves to performance evaluation. The unsupervised learning approach is more of a white-box process as compared to other machine learning approaches like Artificial Neural Networks (ANN) that are black-box processes (Shwartz-Ziv & Tishby, 2017).
- ***Reinforcement Learning*** is characterized by three main concepts. The concept of a cumulative reward for a software agent, an objective or score function that evaluates the reward at any point and the use of feedback to signal whether the choices are towards the optimization of the cumulative rewards. An excellent review of reinforcement learning is found in Sutton & Barto (2018).

An alternative to the popular classification of the approaches to machine learning away from supervised, unsupervised and reinforcement learning paradigm is that proposed in Fürnkranz, Gamberger & Lavrač (2012). The Fürnkranz approach categorizes machine learning into symbolic and statistical approaches. **Symbolic approaches** are characterized as those that involve the inductive learning of symbolic descriptions and are thus examples based and generalization driven. Symbolic approaches include trees, rules and logical representation. On the other hand, **Statistical Approaches** includes statistical or pattern recognition methods like k-Nearest Neighbour (kNN), Instance-Based Learning (IBL), Bayesian Classifiers (BC), Artificial Neural Networks (ANN), Support Vector Machines (SVM) etc.

For data mining to be undertaken, some techniques need to be applied. These techniques can either be viewed as algorithmic or computing techniques. The difference between the algorithmic and computing techniques, as approaches to data mining, is that whereas algorithmic techniques are guaranteed to be finite, computing techniques are not guaranteed to terminate.

Although machine learning is variously defined (Bishop, 2006; Mitchell, 1997; Nilsson, 1996), the overriding concept is the ability of the learning techniques used to not only predict an outcome given some input but also to improve at the task given more experience. Machine learning techniques that could be used in data mining can be classified into either symbolic or statistical-based techniques. Whereas Statistical methods include Logistic Regression, Linear Discriminant Analysis, Bayesian classifiers (Langley & Sage, 1994; Witten & Frank, 2005), Regression (Gunst, 2018; Hardle, 1990), Artificial Neural Networks (Bishop, 2006; Mitchell, 1997) and Support Vector Machines, Symbolic methods include decision trees, rules and logical representations (Mitchell, 1997; Witten & Frank, 2005).

Whereas pure machine learning models like artificial neural networks (ANN) rely on automatic adjustment of parameters in an iterative process and are mainly not transparent, purely statistical methods like Bayesian classifiers rely on noise-free data that follow normal distributions.

The choice of a machine earning methods is guided by suitability for use on some specific data, volumes of available data, required outputs and their levels of transparency and idiosyncrasies of the modelling environment (Anderson, 2007). The machine learning problem, for ease of manipulation, is always essentially reduced to the three concepts of representation, evaluation and optimization.

Representation reduces the machine learning problem to a formal computer language thereby defining the possible classifiers that could be trained on the data. This is in essence, the determination of the hypothesis space, H. Closely following on representation, is the key concept of *Evaluation* that involves the formulation of some performance evaluation function against which the candidate models are evaluated.

Evaluation is, therefore, the basis on which alternative models are weighted against each other. The results of the model evaluation are followed by the related concept of *optimization* that uses the results of the evaluation process to inform the choice of models. Optimization involves the search for the "best" classifier from amongst the candidate classifiers. The best classifier will be that which has the best performance based on the objectively formulated and evaluated score function that could combine multiple measures of performance.

In general, the simplification of the machine learning problem to a model is the very basis of potential misuse and misinterpretation of the process. The process of machine learning, to avoid the pitfall of over-simplification, should follow on the key guiding principles that cover issues around the goal, representation, postulation of assumptions, generalization and handling of the curse of dimensionality, realization of multiple models and simplicity for the practicality of use. These key guiding principles, as documented variously from Bishop (2006), Chao (2011) and Mitchell (1997) are summarized as: -

1. *The Goal:* **-** the ultimate goal of machine learning is to realize models that can be generalized past the observed training examples. Reasonable performance at the model training is thus desired, but must not be the ultimate goal of any machine learning problem.
2. *Representation versus "Learnability":* **-** The ability to represent a problem as a machine learning problem does not directly imply it is learnable. In this context, representability can be viewed as being trivial to the learning task. It is therefore advised that multiple representations for the same problem be investigated.
3. *No Free lunch:* **-** The concept of "No free lunch" implies assumptions about a learning model have to be made to make it generalizable. This is in direct conflict with the reality of the insufficiency of the theoretical underpinnings of machine learning.
4. *The curse of dimensionality* in which a learner is presented with data that has too many features when most of such features are not direct evidence of the

concept to be learnt. The curse of dimensionality calls for a well-considered feature engineering process. The presentation of data with many features not only poses the risk of model overfitting but also a contradiction to the choice of machine learning techniques that are known to be optimal in cases of fewer features and low volumes of data.

5. *Multiple models learning* involve the learning of many models. The multiple models learning approach is encouraged since performances differ with different scenarios and therefore it offers the opportunity of the selection of best performers or the use of results.
6. *Simplicity does not guarantee accuracy* is as opposed to the advocacy by Occam's razor for simple yet predictive models. It is the case that model ensembles outperform single and simple models despite their complexity.

The study furthers the review of two approaches of in machine learning: Artificial Neural Networks (ANN) and Support Vector Machines (SVM) in section 2.8.4 and section 2.8.5 respectively.

2.8.4 A review of Artificial Neural Networks (ANN)

Artificial Neural Networks (ANNs) are a machine learning approach that mimics the interconnectedness of the brain in the modelling process. ANNs have numerous neurons connected to each other to be able to emulate the human brain. The interconnectedness of the neurons is achieved by typically grouping them into layers called input and output layers with one or more hidden layers in between. A typical set with two hidden layers is as shown in Figure 2.12.

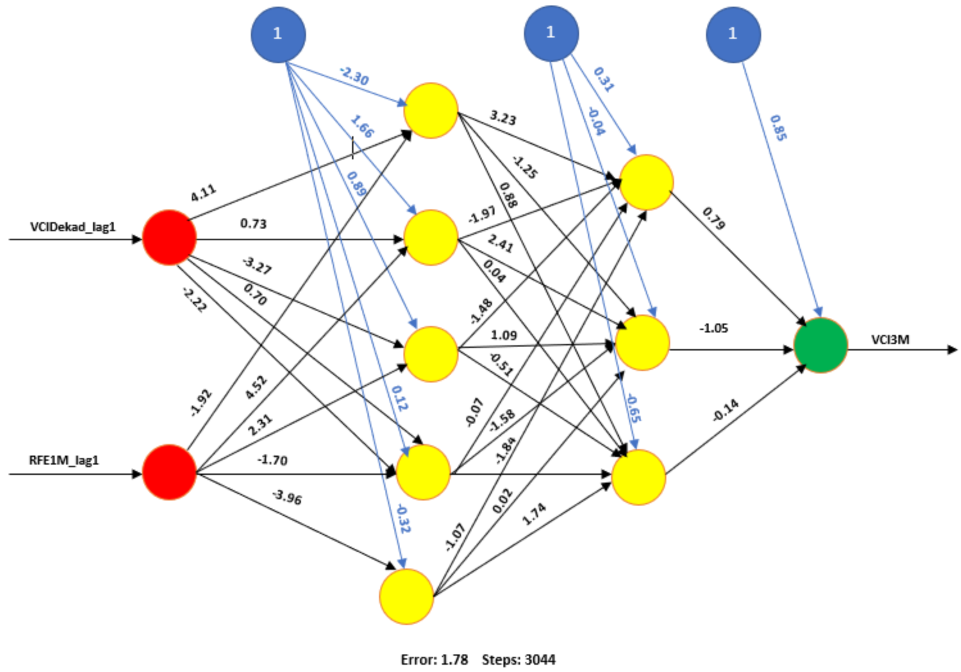


Figure 2.12: An ANN model with an input, output and two hidden layers with the configuration 3-5-3-1.

For regression, the ANN is modelled with the sigmoid unit shown in Figure 2.13 at the centre of the design. In a multilayer network, it is these sigmoid units that are connected in a feed-forward set up as shown in Figure 2.12. The feed-forward network can propagate inputs forward and errors backwards to adjust weights till some predetermined thresholds are met. The algorithm used to feed the inputs forward and the errors backwards till a threshold is met is referred to as the backpropagation algorithm.

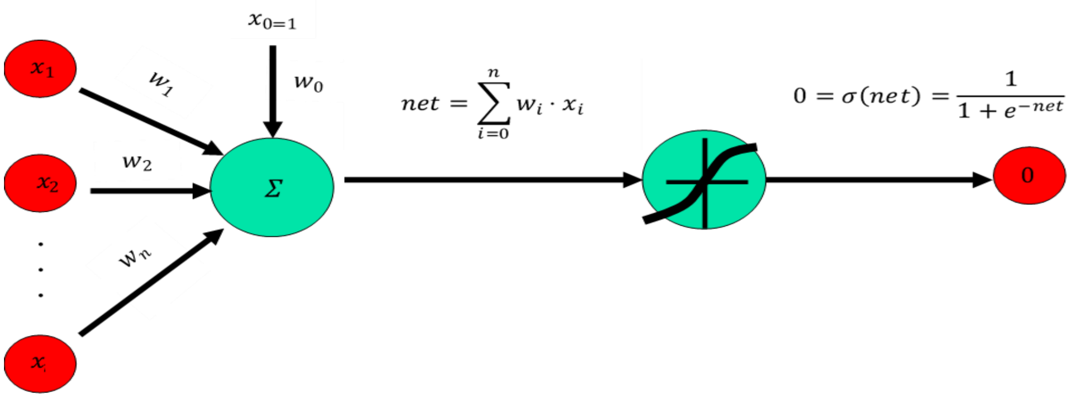


Figure 2.13: The sigmoid unit- the basis for regression modelling using ANN

The sigmoid unit $\sigma(x)$ is acts as the sigmoid function and is expressed as $\frac{1}{1+e^{-net}}$. The sigmoid function has the nice property that the output can be expressed as a function of the inputs such that $\frac{d}{dx}\sigma(x) = \sigma(x)\big(1 - \sigma(x)\big)$. The differential for the expression can thus be obtained without necessarily expanding the expression.

The popularity of the ANN is hinged on its ability to learn discrete, real and vector-valued functions and to vastly remain robust to errors in training data. As documented in Mitchell (1997), ANNs have several characteristics making them suitable for predictive modelling: *(1)* instances can be represented by many attribute-value pairs, *(2)* the target function is either discrete, real or vector-valued, *(3)* training examples may contain errors, and *(4)* long training times are acceptable while faster evaluation is required. The ANNs are, however, susceptible to overfitting (Bilbao & Bilbao, 2017) in which case localised expert models are realised in model training that are not practical for real-life scenarios and hence the models end up with poor performance in test data.

A good description of ANNs is provided in Ramos & Martínez (2013), Bishop (2006), Nilsson (1996) and in Mas & Flores (2008). The study in Mas & Flores (2008) provides a review to the use of the ANN's backpropagation algorithm in remote sensing. On the other hand, the study in Ramos & Martínez (2013) reviews the literature on ANNs and makes a comparative analysis of the performance of different groups of ANN in time series forecasting. The results in Ramos & Martínez (2013) show the multi-layer perceptron (MLP) as the best network in forecasting time series data.

2.8.5 Support Vector Machines (SVM)

At the core of SVMs is the question of separability of data points into unique classes, typically two but in an approach that is extensible to multiple-classes and to linearly inseparable cases. Given that the remote sensing data used for drought monitoring have a sinusoidal trend with the effects of seasonality, it is expected that the examples are not linearly separable. Drought monitoring data is also expected not to follow on any of the standard distributions. We, therefore, review SVMs and their regression implementation as support vector regression (SVRs) as capable of modelling cases of linear inseparability while tolerating noise in the data as documented in Mountrakis, Im & Ogole (2011).

In the simplest approach of the SVM, the key is to create a decision boundary or hyperplane between two classes in a setting that supports the prediction of classes using one or more feature vectors. The aim is to have the hyperplane orientated as to be furthest from the support vectors that are designated to be the closest data points from each of the classes to be separated. The performance of SVM against other approaches has been documented in Wagacha (2003) using empirical evidence while Cristianini & Shawe-Taylor (2000) and Scholkopf, Smola & Bach (2002) provide a comprehensive review of SVMs.

The issues in the use of SVM are documented to include the handling of noise in the data and handling training data that are linearly inseparable. These issues are, however, solved as follows: -

- Noise in the training data for SVMs is handled by having soft as opposed to hard margins. Soft margins allow for misprediction of some of the support vectors. Whereas Figure 2.14 shows the simple case of an SVM without noise, Figure 2.15 shows the effect of noise and thus the need for a soft margin. The case with noise, however, has a cost (C) hyperparameter introduced to take care of the complexity of models. The aim, in this case, is to find the C that neither overfits nor underfits the model.

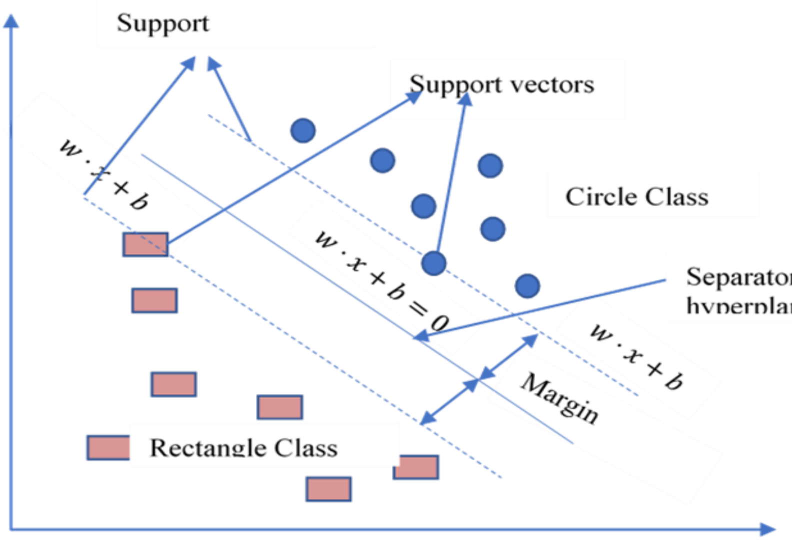


Figure 2.14: Decision boundary of SVM and a hard margin for linearly separable training data

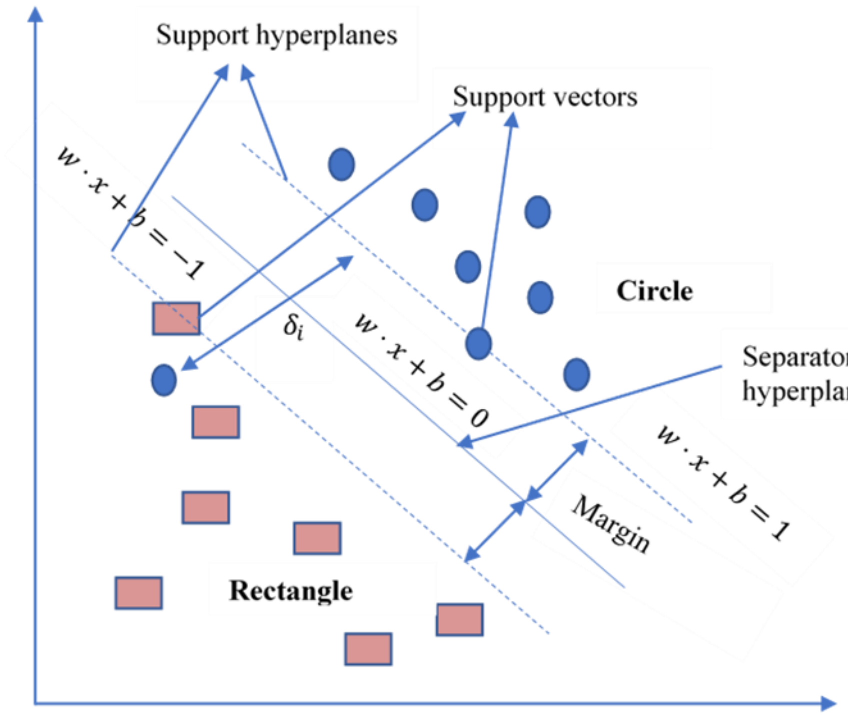


Figure 2.15: Decision boundary and a soft SVM margin with noise accounted for by the cost hyperparameter (C).

- For cases of linearly inseparable training examples, the data is mapped to a higher dimensional feature space that could even be infinite. In some cases, even one step of transformation realizes linear separability on the training data.

This is as illustrated by Figure 2.16 when a transformation from (x_1, x_2) to the expanded (x_1^2, x_2^2) feature space is done.

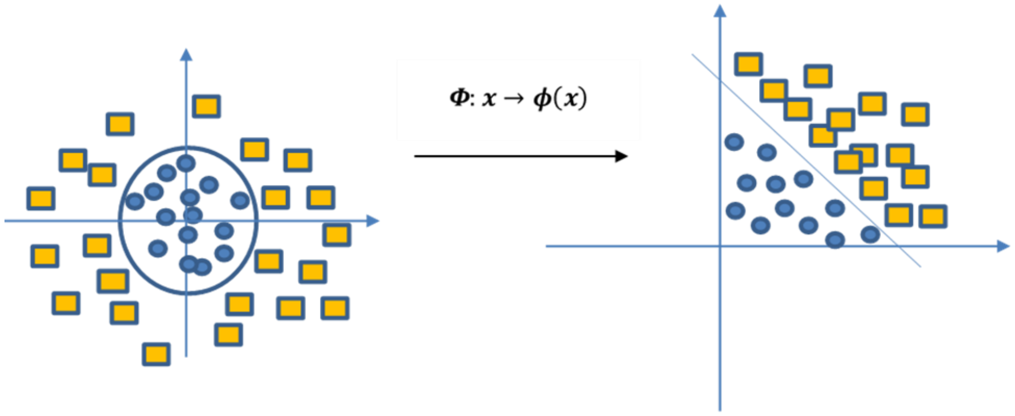


Figure 2.16: Feature transformation to higher dimensionality in an SVM.

Figure 2.16 illustrates the mapping from a feature space to a higher dimensional space to solve the problem of non-linearly separable training data. For the cases where transformation does not yield a linearly separable feature space, a solution can still be formed using Kernel functions. Assuming a 2-dimensional feature space $x(a), x(b)$ the aim is to find the dot product of $x(a.b)$ as is in the linear case of the SVM. However, to achieve this, we make a transformation of the features to a higher dimension $\phi(x_a), \phi(x_b)$ as illustrated in Equation18.

$$\Phi(x_a), \Phi(x_b) \rightarrow \phi(x_a), \phi(x_b) \dots \dots \dots \dots . (18)$$

As documented in literature and illustrated in simplicity in Berwick (2003), there exist some ϕ for which Equation 19 holds implying that the dot product is a function of the inputs. This representation in Equation 19 makes it easy to compute k without expanding ϕ. k is referred to as a Kernel function as it corresponds to the dot product of two feature vectors in some expanded feature space. The kernel function can be any of linear, polynomial, radial basis or sigmoid and it is what is used to model SVMs on non-linearly separable training data.

$$k(x_a, x_b) = \phi(x_a) \cdot \phi(x_b) \dots \dots \dots \dots . \dots . . (19)$$

A special case exists for the application of SVM on regression problems when the output is not a classification but a real value. Like the SVM, the approach is both still non-parametric and permits the modelling on non-linearly separable training data, unlike other approaches that rely on assumptions like Gauss-Markov as is the case for Simple Linear Regression (SLR). Also, like the SVM case, with the error set to some threshold based on the *principle of maximum margin,* the SVR does not care about the prediction for cases where the error (ε)respects the limit. Finally, the SVR also supports the cost parameter and offers high flexibility as it is a distribution-free approach.

2.9 Review of similar projects

2.9.1 Targeted review of past studies

Trnka et al., (2018) did a survey on the priority areas of drought research amongst experts from different backgrounds. 65 experts from 20 different professions across 21 nationalities document the priority areas to include drought forecasting and prediction amongst others like drought monitoring, impacts mitigation through both drought resilience and adaptation to drought. The best drought predictions are then proposed to be best those done at different time scales and also that incorporate multiple ensembles of numeric weather models.

Apart from the survey of experts in Trnka et al., (2018), there is the presentation of a comprehensive review of methods in Mishra & Desai (2006) that are used in predictive drought monitoring. The limitation of stochastic approaches like autoregressive moving average (ARMA) model and their generalizations of the autoregressive integrated moving average (ARIMA) and their seasonal applications SARIMA are documented. ARIMA models are also reviewed in Belayneh & Adamowski (2012) and Mishra & Singh (2011) with their limitations described in two-fold. First is their inability to overcome the random nature of droughts and second is their poor performance in modelling non-linear and complex events like drought as also elaborated in Agana & Homaifar (2017).

The approaches in drought prediction are classified into five broad categories in Mishra & Singh (2011). These classes are *Regression analysis*, *Time series analysis* (ARIMA based), *probability models*, *neural networks* and *hybrid models*. It is the take of Mishra

& Singh that there are better prospects to the use of hybrid models that incorporate climatic indices in the prediction of complex events like drought.

We discuss the past studies in light of the above key concepts presented in the reviewed papers. The studies expected to overlap the categories used in the review due to inherent complexities in drought studies. The categories include:

- Types of variables used and if entire drought types are covered and hence the integration of socio-economic data
- Nature of study based on whether they are downscaling approaches or investigation of new approaches.
- The type of models used and whether they are regression, probabilistic, neural networks, time series or hybrid models.

Review by types of variables and types of drought

Most of the studies reviewed are either single variable/indicator index models or cover a single type of drought from the four types of meteorological, hydrological, agricultural and socio-economic. Most studies document the use of remote sensing data as compared to socio-economic data that is indeed rarely used across the studies. The non-use of socio-economic data implies that the quantification of the impacts of drought is therefore largely missing from literature. The remote sensing data used can be classified into meteorological, hydrological and hence water balance or agricultural.

Belayneh & Adamowski (2012) uses precipitation data transformed into standardized precipitation index (SPI) with the data aggregated over 1, 3 and 12 months to predict exclusively meteorological drought. A similar approach of using only one precipitation-based variable or index in the prediction of one type of drought also includes that documented in Wetterhall et al. (2015). The aim of the study in Wetterhall et al. (2015) was to predict the probability of dry spells and below normal precipitation over a season with lead times of between 0 and 4 months. Other studies that use SPI as a single variable to predict meteorological drought include Khadr (2016) as well as Wichitarapongsakun et al. (2016). The use of a single variable/indicator index is also found in Ali et al. (2017) and Le et al. (2016) that uses Standardized Precipitation

Evapotranspiration Index (SPEI), an extension of SPI and potential evapotranspiration (PET) in the prediction of meteorological drought.

Despite the tendency to use a single variable to basically predict one type of drought, there are a few approaches that deploy more than one variable to either predict on type of drought or to predict two of the drought types. Such studies include Morid, Smakhtin & Bagherzadeh (2007) that together with Huang et al. (2016) use both SPI and Effective Drought Index (EDI) to investigate the severity, duration and extents of drought events. The comparison between EDI and SPI documents the EDI to outperform the SPI in both Morid, Smakhtin & Bagherzadeh (2007) and Huang et al. (2016). The use of SPI in conjunction with another variable is additionally documented in Yuan et al. (2017) that uses SPI as the predictor variable together with standardized streamflow index (SSI) to predict hydrological drought conditions.

Apart from the studies above that document use of remote sensing data, there is a set of studies that either advocate for the use of socio-economic data in drought monitoring or those that proceed to incorporate socio-economic data in drought monitoring. These socio-economic data studies are summarized as follows:

- ***Hao, Singh & Xia (2018)*** in reviewing the advances, challenges and future prospects in the prediction of seasonal droughts point out the limited existence of studies that document the prediction of drought effects. The study observes the availability of documentation on the prediction of drought signals using different remote sensing indicators. The study points out that the identification of drought signals does not come with the identification of the effects of the droughts on society. The authors advocate for the exploration of indicators appropriate for the quantification and prediction of the effects of drought in addition to the systems that monitor the drought signals.
- ***Bachmair et al. (2016)*** document the lack of "ground-truthing" of drought monitoring variables as viewed in terms of ensuring such indicators represent local drought conditions and/or their impacts. The review and survey study observes the overspecialization in agricultural drought monitoring at the

expense of other drought types and also the analysis of drought impacts in term of impacts on vegetation. The survey study had responses from 33 DEWS experts and it advocates for the inclusion of the environment and society in drought monitoring. The study refers to the inclusion of socio-economic impact data as the "missing piece" in drought monitoring. The possible challenges to use of socio-economic data as outlined in the study are: *(1)* cost of collection of impact data *(2)* the many possible impact indicators; *(3)* differences on drought understanding and perception of drought impacts; *(4)* the interaction between impacts and vulnerability of people; *(5)* the delayed response between droughts and their impacts and; *(6)* the complexity of multi-causality of impacts.

- ***Jenkins (2012)*** presents the analysis of economic and social impacts of droughts within future projections of climate change. The analysis includes both direct economic drought costs and social drought effects. The most notable things about the study are that it is both ex-ante and is predictive. It analyses past droughts and makes predictions of impacts of future droughts. Direct impacts are identified based on the model used in Hochrainer et al. (2007).
- ***Massarutto et al. (2013)*** present an ex-post analysis of the socio-economic and environmental impacts of historical drought events. The analysis is however restricted to the agriculture and power sectors using the consumer surplus theory.
- ***Musolino, Massarutto & De Carli (2015)*** focused on a purely agricultural market focusing on whether socio-economic impacts of droughts produce winners as opposed to earlier approaches that focused only on losers. The study focused on the drought impacts on market prices.
- ***Enenkel et al. (2015)*** explore how to integrate non-environmental information sourced via smartphones to augment agricultural drought monitoring in the context where future uncertainties in drought prediction are understood. The main drive in the study was to find, out of collaboration, better ways to turn

data streams into useful information for decision support. A framework is proposed for an operational decision support system that includes a mobile-driven collection of socio-economic information. The study does not, however, provide a framework for the operationalization of the integration of socio-economic data or how such a system would look like in the monitoring or prediction of future droughts. Possible data suggested by the study include access to potable water, the level of malnutrition and the current prevalence of diseases.

- ***Enenkel et al. (2016)*** make two key observations. First is the use of geospatial information as almost indispensable in decision making as pertains natural disasters. Second is the need for technology transfer especially between research and application-driven by the need for user-friendly tools that couple drought risk to socio-economic vulnerability. The study proposed a super-index- the enhanced combined drought index (ECDI) that integrates rainfall, soil moisture, land surface temperature and vegetation vigour. The operationalization of ECDI has integration with data on socio-economic impacts of drought collected using mobile phones. The predicted future ECDI values do not, however, integrate the socio-economic data and neither are the socio-economic impacts predicted.

Review by Type of study

A review of the past studies can see the studies categorized as either scaled-down versions of global drought monitoring approaches or as the application of known computing methods in novel ways. Studies that scale down global models for drought prediction are documented in Wetterhall et al. (2015), Trambauer et al. (2015), Huang et al. (2016), Yuan et al. (2017), Turco et al. (2017) and Štěpánek (2018). This approach involves taking an existing model and using it in the context of a new set up.

In Huang et al. (2016), the SPI and Effective Drought Index (EDI) are used to assess severity, duration and spatial extent of drought events using the Statistical DownScaling Model (SDSM) approach. A similar downscaling approach is used in

Wetterhall et al. (2015) following on the European Centre for Medium-Range Weather Forecasts (ECMWF) seasonal forecasts system 4 (SYS4) to make seasonal forecasts of dry spells over the Limpopo basin during the rainy season with lead times from 0 to 4 months. The Limpopo basin also saw the use of three downscaled global models in Trambauer et al. (2015) in the forecast of drought.

The ensemble streamflow prediction system (ESP) is compared to the operational dynamical forecast system ECMWF seasonal forecast- System 4 (SYS4) in predicting summer drought in Europe (Turco et al., 2017). Most interesting is the ensemble of downscaled models found in Štěpánek (2018) that uses 5 global numerical weather prediction (NWP) models including ECMWF and two soil models in an ensemble to forecast soil moisture and drought intensity. Across these studies, the most interesting is the interplay between model downscaling and the opportunity for ensembling. A further review of the application of down-scaling is, for example, provided in Wilby & Dawson (2013).

The greatest limitation of the downscaling approach is in the definition of drought event and the follow-up variables that are then used for the monitoring. Droughts as earlier reviewed in section 2.3 on the “Key concepts in the definition of drought” was noted to have both spatio-temporal attributes. The non-universality in the definition of droughts would make for non-applicability of some of the data in some spatial extents. Additionally, given that the scaled-down models have to use variables similar to those used in the global versions creates a shortcoming especially in the case where a specific kind of variable turns out non-responsive to droughts in a given spatial extent or even when the drought type of interest changes from the focus in the global models. The studies that defined any of the computing methodologies afresh by learning new models are described in the next section on the review by methods of drought prediction.

Review by methods of drought prediction

The methods of drought prediction, apart from those that use scaled-down versions of existing global models and that involve the use of novel computing methods are documented in quite a sizeable number of studies. The studies are reviewed based on whether the deployment is made of a single method or is of multiple methods. The single and hence pure methods include any of artificial neural networks (ANN), multiple linear regression (MLR), Hidden Markov Models (HMM), random forests (RF), time series prediction (TSP), Bayesian frameworks/ networks (BN), autoregressive integrated moving average (ARIMA). Multiple models, on the other hand, entail an assembly of two or more of these specific methods. A review of the studies by the methods is provided here next.

Pure artificial neural networks are used in amongst others: Morid, Smakhtin & Bagherzadeh (2007), Maca & Pech (2016), Le et al. (2016) and Ali et al. (2017). The following is an in-depth review of each the drought prediction studies with respect to their contributions and limitations.

- ***Morid, Smakhtin & Bagherzadeh (2007)*** used an artificial neural network to formulate a continuous function of rainfall that predicts drought. It uses Effective drought Index (EDI) and the SPI in addition to climatic indexes Southern Oscillation Index (SOI) and North Atlantic Oscillation (NAO) index. An R^2 of 0.66-0.79 is reported. EDI models are also reported to outperform SPI models in this study. The documentation indicates the superiority of the EDI forecasts over those of the SPI. The predictions are documented to only indicate a possibility of future dry or wet conditions as opposed to quantifying drought severity. One limitation of this study is the avoidance of synergies from the two datasets used in the prediction of droughts.
- ***Masinde (2013)*** uses the artificial neural networks (ANN) and the Effective Drought Index (EDI) to solve the twin problems of provision of short- and long-term drought forecasts and the specification of the severity of the drought. The approach, however, covers meteorological drought and perhaps early phases of hydrological drought.
- ***Maca & Pech (2016)*** documents the importance of time series indices in drought monitoring. The study has quite a set of interesting facets. The study uses ANNs in an integrated formation and compares their performance with

the known Multilayer Perceptron (MLPs). The approach used five different objective optimization functions to realized four ANN models that were then used in a hybrid formation to predict the values for SPEI and SPI. This study, even though it does not explicitly call the approach ensembling is in our opinion a homogeneous ensemble approach that uses purely ANN models with a supermodel for the ensemble. The limitation in this study, however, lies in the fact that the combination of the members into the ensemble not investigated. In our opinion, apart from the hybrid approach that is definable as an ensemble in our opinion, this approach still predicted only meteorological and partly agricultural drought without the integration of agricultural and socioeconomic drought indicators. Strictly speaking, the model in this study does not define drought events and so remains a system for the forecasting of future SPI and SPEI values rather than drought.

- ***Le et al. (2016)*** do index forecasting using lagged values of the same index. The study also documents the use of climate signals as predictors of continued interest. The study uses multiple SPEI time steps of 1, 2 and 3 months and also uses lagged climate variables to predict future SPEI conditions using multilayer perceptron feedforward neural networks due to their popularity. Five years of data covering the period 2001-2006 was used in the study. The study period was documented to have had one drought event and the aim was to select the best performer ANN model for a single extent that best predicted this drought. In our opinion, there are two distinct limitations to this approach. First, the period of study amounts to a short period of time with only one drought episode. Drawing references from such a short period for generalization into the future could be a little shaky for lack of enough instances to achieve full model calibration. Though ANNs handle scarcity of data, we opine that this is quite a short period for weather-based events to be well understood by a model and subsequently predicted. Second is the use of one extent with the same dynamics recorded. Generalization beyond the extent of learning would be a tall order in our opinion.

- ***Ali et al. (2017)*** document the popularity of the use of drought indices as the tools for drought management. This study uses the Standardized Precipitation Evapotranspiration Index (SPEI) that is an extension of both precipitation and potential evapotranspiration (PET). The use of ANN with a feed-forward topology with backpropagation learning was deployed. Three (3) layers with the configuration 30-8-1 were shown by experimentation to be the best performer with 30 past values of each index used. MAE, R^2 and RMSE were used as performance measures. Correlation coefficients of between 0.887 to 0.987 for a 1-time step prediction were reported. This implies a determinant of correlation, R^2 of 0.79 that represents very good performance. It is, however, to be noted that the study had multiple models for the different spatial extents for the good return in performance. The development of a model for each of the different regions of interest is bound to make this approach difficult to scale for multiple extents.

Other pure methods include Multiple Linear Regression like in Huang et al. (2016) that is used in the context of a scaled-down model, Hidden Markov Model HMM) in Khadr (2016) that used SPI to predict meteorological drought. Remarkably the study in Khadr (2016) posts an R^2 of 0.96 and RMSE of 0.20 is reported for the study for meteorological drought and for 1 time-step lead time. This, in our opinion, is quite some admirable prediction for a 1-unit time step. The dip in performance using this approach for the next time step to R^2 below 0.56 however make it quite declining in performance with increased lag. The use of random forest (RF) is advanced in Shah et. al (2017) that used multiple indicators including temperature, precipitation, evapotranspiration to predict what amounts to both meteorological and hydrological drought leaving out both agricultural and socio-economic aspects of drought. Time series prediction and Bayesian approaches are also documented to have been used in drought prediction including but not limited to the studies in Wichitarapongsakun et al. (2016) and Madadgar & Moradkhani (2013) respectively.

2.9.2 Multi-model method studies

Increasingly, prediction of droughts using remote sensing and socio-economic data are adopting the use of multiple models as has been the case in climate-based predictions at a global scale. The models are either in comparative studies or those that proceed to use the opportunities offered by the multiplicity of the models in the actual predictions. These approaches have been documented in Moustris et al. (2012), Belayneh & Adamowski (2012), Mishra & Desai (2006), Agana & Homaifar (2017) and Rhee & Yang (2018). These studies are summarized as follows: -

- ***Moustris et al. (2012)*** though not on drought monitoring is one that uses climate-based data to do a surface ozone forecast using a Multiple Linear Regression (MLR) model against ANN. The results indicate the competitiveness of the results with an R^2 0.65 to 0.67 in favour of the ANN. This study is one basis for the indication of the superiority of the ANN to most of the statistical methods. The use of similar variables across the two models for comparability of results in this study is in our opinion well informed.
- ***Belayneh & Adamowski (2012)*** used the SPI aggregated over 1, 3 and 12 months to investigate the appropriates in the prediction meteorological drought. Predictions were done using wavelet neural networks (WN) ANN and Support Vector Regression with results compared to ARIMA. The WN is shown to produce the most effective model for the forecast of conditions with 3-month and 12-month lead times. As expected, the study documented the techniques as all degrading in performance with increased lead times in the predictions.
- ***Madadgar & Moradkhani (2013)*** does a seasonal forecast of drought as defined by Standardized Streamflow Index (SSI). The SSI is defined akin to the SPI. The focus of this study is the prediction of hydrological drought and the method used is the Bayesian framework method in the context of a multivariate probabilistic framework. First-order and second-order conditional probabilities were calculated for the subsequent seasons.

- ***Lu et al. (2014)*** used a combination of fuzzy logic and neural networks in an improved neural network fuzzy inference model to predict weather conditions such as precipitation. The study documents the superiority of the results of the fuzzy networks as compared to purely ANN models. The model, however, only made predictions of meteorological droughts and does not use drought effects data or even vegetation data to model the other types of drought. All data used is from a single source and thus there is no evaluation of multiple sources of data.
- ***Agana & Homaifar (2017)*** document the need for more complex models to overcome the random and non-linear nature of droughts that are made worse by climate change. The suitability of ANNs with two hidden layers is documented to overcome the problem of non-convex optimization. Using a Deep Belief Network (DBN) with 2 restricted Boltzmann Machines, drought predictions are made using lagged values of Standardized Streamflow Index (SSI). The study uses the 2-dataset split approach in the development of the models whilst using RMSE and MAE as measures of performance.
- ***Rhee & Yang (2018)*** uses Extra-Trees (EXT) and Adaboost for classification with the 80:20 % data split approach. The Machine Learning Approaches of EXT and Adaboost are documented have outperformed bias-adjusted forecasts. Despite the development of multiple models by the study, an opportunity was lost in this approach to investigate the effect of bettering prediction through model ensembling and possibly the integration of socio-economic data.

2.10 Emerging trends in drought monitoring

We discuss the emerging trends based on the review of past studies as angled from the noted trends in research in addressing the documented limitations of the current approaches. We cover trends in both data and methodology.

First, the studies reviewed above show the limitations of stochastic approaches to the prediction of droughts. The limitations of the popularly used ARIMA/SARIMA models are discussed based on their inability to model both non-linearity and nonstationarity that droughts suffer as a result of external impacts like those arising from climate change. The alternative approaches to ARMA based models, however, together with the availability of drought indices over long term periods offer an opportunity for robust and reliable predictions of future drought events.

The second observation is that it is increasingly the case that multiple indices are being used in drought prediction. Most of the studies either use one variable or aim to predict one or two stages of drought with meteorological and hydrological drought being the most commonly studied.

The third is the emergence of the need for the use of socio-economic data in drought monitoring as advocated for in the studies earlier reviewed and particularly in Enenkel et al. (2015). Currently, consideration of impacts of drought is generally neglected by most drought monitoring initiatives as observed in Bachmair, Kohn & Stahl (2015) and in Wilhite, Svoboda & Hayes (2007). The use of socio-economic data would lead not only to the prediction of drought severity based on satellite-derived data but also the possibility of quantifying and predicting the possible impacts of such drought should they occur. This would make for operational systems that measure both drought severity and drought impacts based on data sources from field-based observations as is the case in Enenkel et al. (2016). The use of Socio-Economic Data (SED) in drought monitoring is increasing getting focus because of the need to need to assess the effect of droughts on communities. The ultimate goal of any drought monitoring initiative should be towards keeping the communities protected from the resulting economic shocks through drought mitigation. Drought mitigation should stem from a well-structured drought preparedness process that is informed by identifying pre-planned activities in the non-drought phase through a formal process referred to as drought contingency planning. According to Enenkel et al. (2015) and the Enenkel model in Figure 2.17, there is, therefore, need for drought monitoring systems that incorporate remote-sensed weather-based data, socio-economic data and elements of forecasting into drought early warning systems (DEWS). In this approach, remote sensing data would provide the objectivity of defining exposure to drought and its severity. The socio-economic data would then attempt to quantify the effects of drought while forecasting would be used to provide a lead time view of the progression in both the drought and the drought effects. The operational decision support system (ODSS) would, therefore, be a computer-based application that collects and analyses drought-related data to facilitate both accelerated and quality decision making for drought management and drought planning.

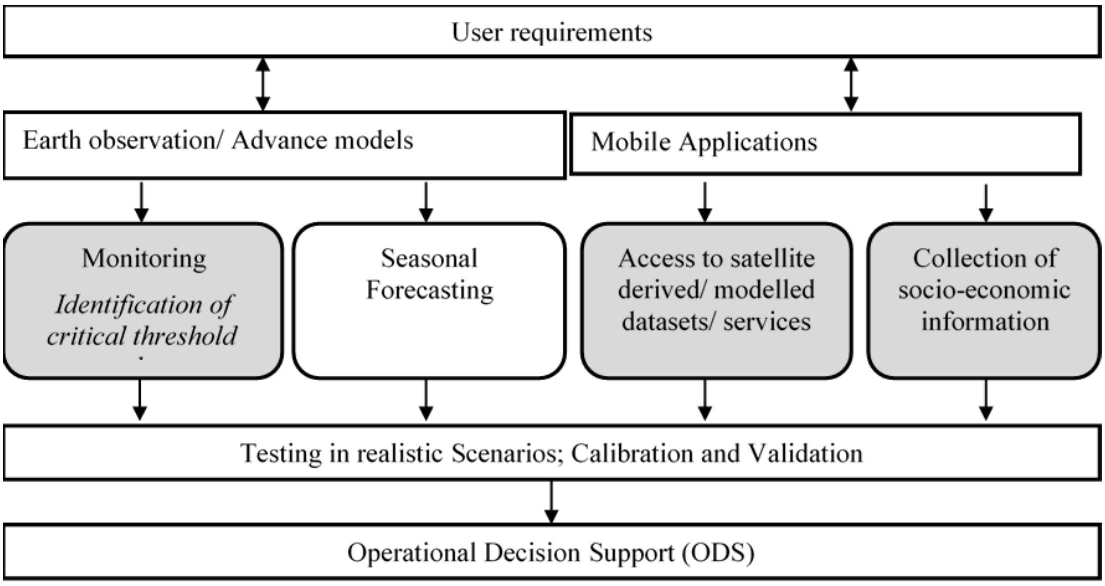


Figure 2.17: The Enenkel proposed framework for an operational decision-support system (ODSS).

Fourth is the demand for approaches that utilize more than one modelling methodology in the prediction of drought. These approaches have different setups ranging from comparative studies that then inform the choice of methods to those that aim at improving the performance of predictive models. The most desirable of the usage of these multiple models would be in a setup where they are used to offer predictions which would then be used for integration to obtain a single possibly stable model offering better performance in prediction. Such would be an ensembling modelling approach.

2.11 Model ensembling as a trend in predictive studies

2.11.1 Definition of ensemble learning

While Zhou (2012) defines ensemble learning as a machine learning paradigm that trains multiple leaners with a view to solving the same problem; Re & Valentini (2012) characterize it as an approach that efforts to mirror the human approach to decision making through the combination of expert opinions. The common thread in the definition of ensemble learning is centred around the combination of different predictors to produce a single prediction or classification as it were.

Model ensembling, in summary, is therefore defined as the approach to prediction-based studies in which multiple models are built whose individual decisions are then combined in some way to make estimations of target variables.

2.11.2 The aim and reasons for model ensembling

The aim in ensemble modelling is documented to be the building of not very predictive models and subsequently transforming them into super classifiers without necessarily generating any new powerful algorithms. An alternative reference to this learning paradigm is meta-modelling or meta-learning and is most appropriate for weak learners.

The basis for model ensembling is both theoretical and practical with the most common ensembling approaches designated as bagging and boosting. The use of bagging and boosting has been documented earlier in Quinlan (1996) and in Drucker & Cortes (1996) due to their effective use in decision trees. Opitz and Maclin (1999) have documented the tendency for experts to use model ensembling in decision tree learning as a result of the faster speeds of learning that characterize them. This explains the rarity in the use of the model ensembling approach for the slow training approaches like artificial neural networks.

The justification for model ensembling is documented to include more accurate predictions compared to individual members and better performance in generalization as a consequence of possibly model specializations in the minimization of different errors on training data. Illustratively, multiple classifiers only get a wrong ensemble classification when more than half the base classifiers are wrong which will be at a reduced probability as compared to a single classifier.

2.11.3 Key Issues in model ensembling

Apart from the question of determination of the members of an ensemble, the other key concepts in model ensembling include the trio concepts of bias, variance and the variance-bias decomposition. While the variance of a predictor is an indication of its sensitivity to small differences in the training set, bias is a function of error in the assumptions in the model. The two concepts are visualized in Figure 2.18.

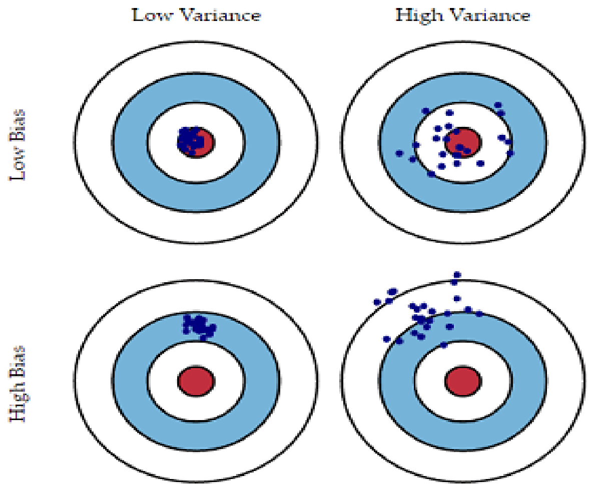


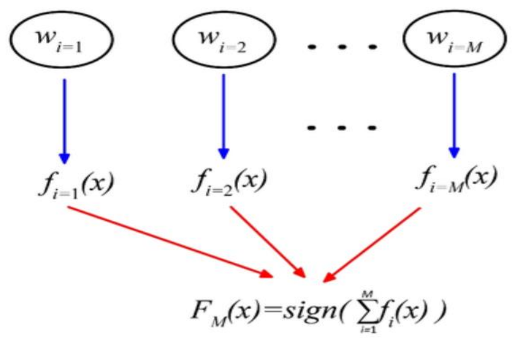
Figure 2.18: Visualization of bias-variance trade-off (Fortmann-Roe, 2012).

Bias in models is related to the concept of underfitting, in which case high bias models are those built so simple that they perform poorly in predictions compared to reality despite exposure to different datasets. On the other hand, variance measures the change in model performance with a change in training data used. Low variance is desirable since it implies model performance does not change with changes in datasets. In modelling, the two concepts have an inverse relationship. The desired point is the low bias- low variance disposition. Since the formulation of the underlying target function is not known, the bias-variance trade-off is key to realizing a good prediction performance. Generally, variance denotes sensitivity to small fluctuations in the training dataset while bias is a result of erroneous assumptions in the model that makes them perform poorly while trade-off between the two is also referred to as the Variance- Bias decomposition. It is the expectation that model ensembling does reduce both bias and variance to lead to the achievements of the low bias-low variance point for best model performance.

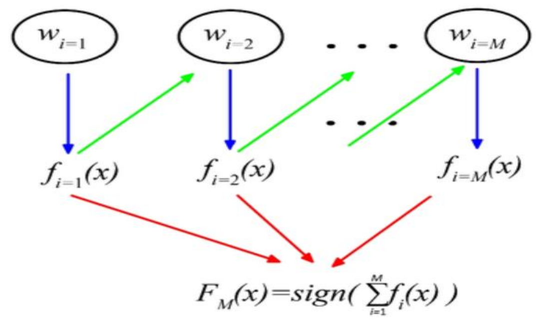
One error that affects model performance as described in Zhou (2012) is the concept of intrinsic noise that demarcates the lower bound on the expected error of any learning approach on the target. This is based on the underlying function being approximated not being well understood in most cases.

2.11.4 Methods of model ensembling

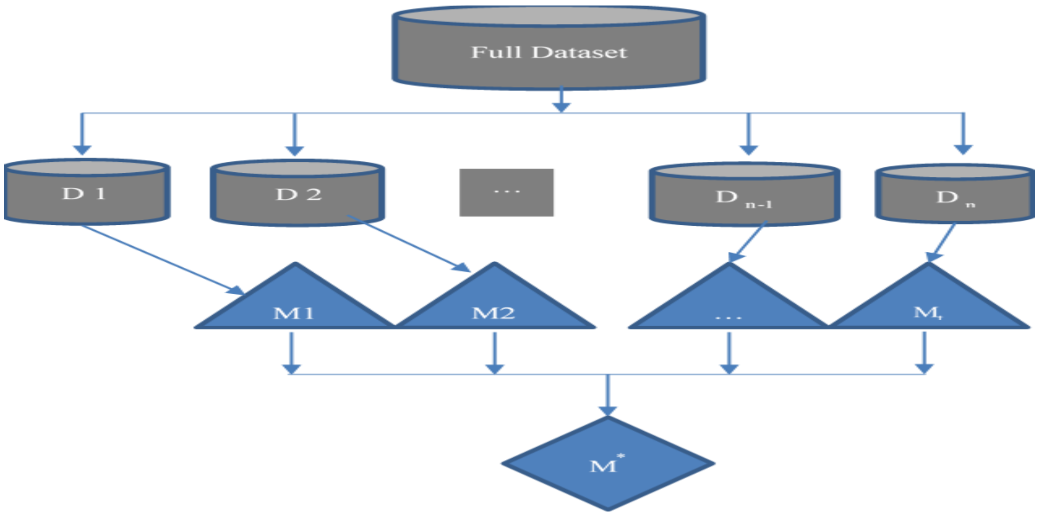
The three methods of model ensembling are variously described as bagging, boosting and stacking. The three approaches are as depicted in Figures 2.19 (a)-(c).



(a) Bagging



(b) Boosting



(c) Stacking

Figure 2.19: Methods of model ensembling: (a)bagging, (b) boosting and (c) stacking.

Figure 2.19 shows the model ensembling methods:

- Bagging (a) in which - $w_{i \dots} w_M$ represent the M bootstrap samples in which the different M learners from the same base algorithm are built and their outputs combined using the aggregation function $F_M(x)$.
- Boosting (b) that has the learners trained sequentially on the poor performing cases.
- Stacking (c) that has multiple initial learners ($M_1 \dots M_n$) developed from potentially different bootstrap samples ($D_1 \dots D_n$) and potentially different algorithms. The outputs of the initial multiple learners are used as inputs to a meta-learner from which outputs are derived.

In Bagging, multiple bootstrap samples are derived from the training data randomly and with replacement. Different learners of the same base algorithm are then trained on the different samples. The outputs of the learners on the different samples are then combined to realize a single prediction by for example averaging the individual predictions. As opposed to bagging, boosting manipulates the training set but the leaners are trained sequentially on poorly predicted cases. The trained models are then used as a committee whose outputs are weighted. The key difference is that bagging derives the models in parallel while boosting derives the models sequentially with iteratively re-weighted training examples. Boosting is thus susceptible to model overfitting.

Stacking, as shown in Figure 12.9(c) and also documented in Džeroski & Ženko (2004), is an approach that builds multiple first-level learners, potentially by using the same algorithm or different learning algorithms. The initial level learners are subsequently combined by a second level learner that is referred to as a meta-learner. The ensembles are thus either homogeneous when the same technique is used to build all the base models or heterogeneous when the base models are borne out of different algorithms. We infer the heterogeneous ensemble as the model ensemble in its truest form.

2.12 Conceptual Framework

A conceptual framework is variously defined but with generally no agreement in its standard definition. Jabareen (2009) defines the Conceptual Framework as a network of interlinked concepts that together provide a comprehensive understanding of a phenomenon or phenomena. The conceptual framework, therefore, provides a general representation of the relationships between things in a given phenomenon. More importantly, conceptual frameworks help in the communication of the ideas out of research (Ivey, 2015).

The conceptual framework can be viewed in terms of the two ways in which it shapes the research process. First, the conceptual framework is used to define the research and to outline courses of action based on a preferred approach to the investigation of an expressed problem. Second is the use of the conceptual framework as an organized way of thinking about how and why a project takes place and consequently on the ways to ensure it is easy to communicate the output of the research to others.

2.12.1 The Elements and Concepts

The overall objective of this study is to investigate the use of model ensembles in the prediction of both drought and drought effects using remote sensing and socio-economic data. The tasks involve the identification of drought monitoring variables based on a broad definition of drought based on all the four types and the investigation on how the variables are related. Subsequently, we use the variables to build ensembles of both homogeneous and heterogeneous methods to predict both drought severity and drought effects. The key concept is drought as measured by different indicators based on the types of drought.

The investigation of the concept of drought and its prediction is based on a predictive research model that has a series of experimental models that are developed and run against the observational data, both remote sensed and socio-economic in nature.

2.12.2 The Conceptual Framework for the study

The conceptual framework for this study is developed based on the key concepts derived from literature review. These concepts are then woven together to define a framework that facilitates their review as guided by the formulated objectives and research questions. These concepts are as briefly outlined below:

- The definition of drought severity as a continuous measure that is to be predicted based on variables derived from the review of literature. This is the first variable to be predicted. Drought severity is indicated by the VCI3M values and its prediction based entirely on remote sensing data.
- The definition of drought impact on nutrition as a measure of the effects of the drought that also depends on drought severity. This is the second variable to be predicted.
- The concept of progression of drought based on types of drought as measured by different indicators. This gives rise to the opportunity to group the indicators by types of drought.
- The intention to use multiple modelling methodologies in an ensemble approach to predict the two variables on drought severity and drought impacts on nutrition for children under five years.

The conceptual framework as shown in Figure 2.20 outlines the relationships between the predicted and predictor variables. Variables are categorized based on the type of drought they are typically used to monitor.

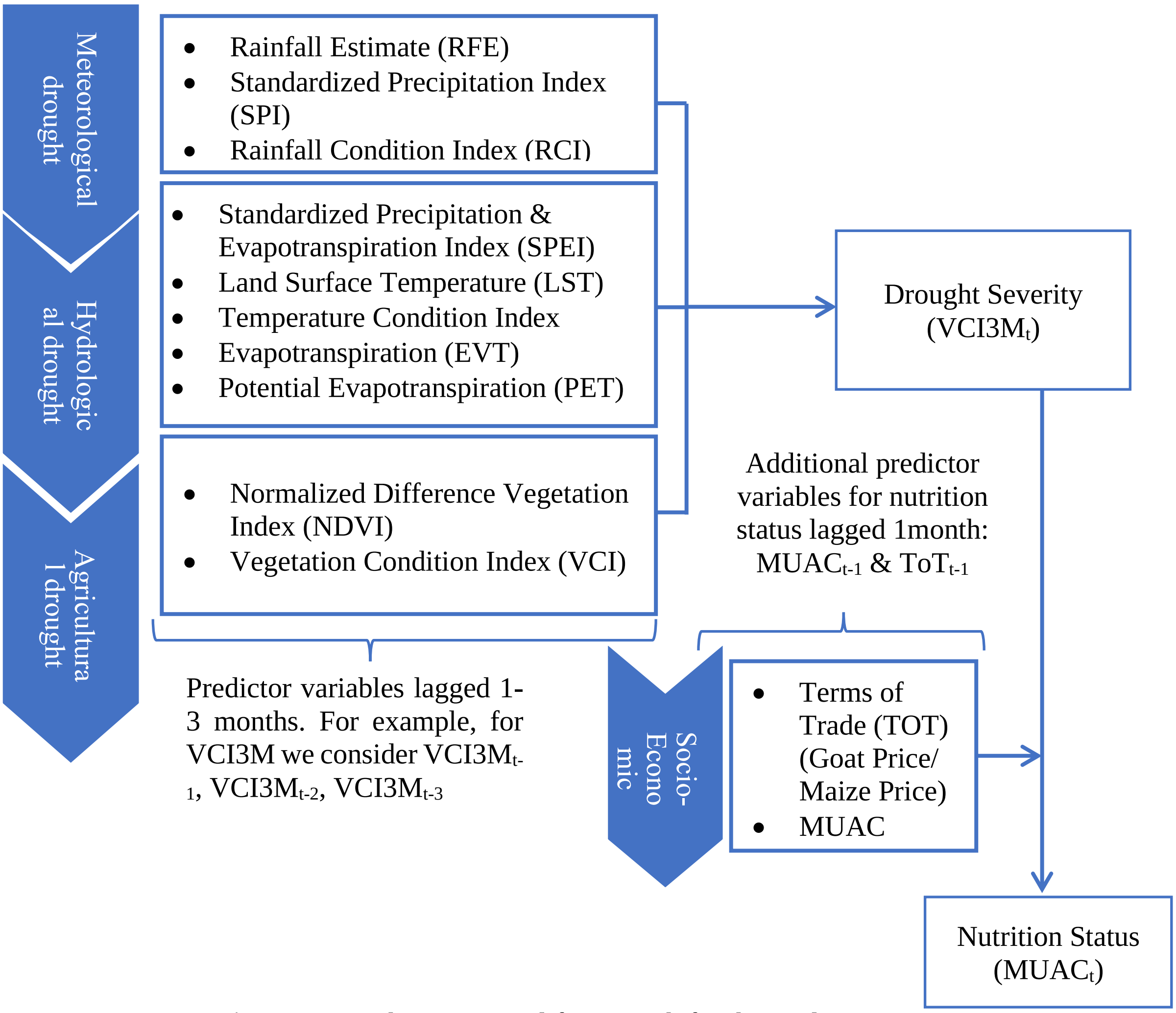


Figure 2.20: The conceptual framework for the study

The conceptual framework identifies the grouping of the predictor variables and the process of their deployment to predict the dependent variables that define both the drought severity and drought effects on nutrition. The predictor variables are values at time *t-1* to time *t-3* months while the socio-economic variable values are at time *t-1*. For each instance, the predicted values are those at time *t*.

From the conceptual framework, given that we use time series data at a monthly frequency, the operationalization of the model will see two district parts of the whole system. These district parts are:

- The monitoring portion of the framework which is characterized by the calculation of the monthly drought indexes from both remote sensing and socio-economic data. This forms the monitoring part of the system. Given we lag the variables in the model, this is equivalent to the t-1 values of the variables.

- The prediction of the future drought severity and drought effects as indicated by nutrition conditions at time t is the predictive/ forecasting component of the framework.

The framework, therefore, gives a predictive approach to drought monitoring. In operationalization, not only do we monitor drought conditions but we also make predictions of future conditions for drought severity and drought effects. This approach can be referred to as predictive drought monitoring in the context of this study.

The first prediction task for the prediction of drought severity is accomplished by identifying multiple models judged to be predictive of drought severity. The identified drought severity models are themselves used to predict drought effects. In addition, the recombination of the drought severity models with the lagged values of socio-economic variables are also used in the prediction of drought effects.

Though the terminologies of dependent and independent variables are interchangeably used with predicted/target and predictor variables respectively, in this study we adopt the latter that is most appropriate for correlational studies due to the absence of treatments and control groups as is the case in experimental studies.

2.13 Literature Review Summary

The gaps identified in the literature review can be summarized into the following major issues:

2.13.1 Prediction of drought signals without recourse to drought effects

Prediction of drought signals and future drought conditions is majorly done using biophysical indicators but with limited investments in the prediction of drought effects (Hao, Singh & Xia, 2018). The studies that advocate for drought early warning systems (DEWS) that incorporate socio-economic data on drought effects include but are not limited to Jenkins (2012), Massarutto et al. (2013), Musolino, Massarutto & De Carli

(2015), Enenkel (2015) and Enenkel et al. (2016). Garrido (2014) for example advocate for monitoring effects on demand and supply markets while Jenkins (2012) expounds on the Hochrainer model (Hochrainer et al., 2007). Massarutto et al. (2013) advise on the quantification of the socio-economic effects of drought events and the modelling of the relationship between variables that monitor drought severity and the subsequent socio-economic impacts associated with drought.

These studies that propose the incorporation of socio-economic data in drought monitoring do not offer how these would be integrated into operational drought monitoring systems and by extension in the prediction of droughts. We, therefore, recognize an opportunity in the furtherance of the modelling of drought effects using socio-economic data.

2.13.2 Univariate over Multivariate systems

Most drought monitoring and drought prediction systems are built around the use of a single variable in either the prediction of droughts or even a single variable in the monitoring of droughts. The use of the single variable is despite the existence of many variables and many data sources even for the same variable. The univariate systems are popular despite the existence of multiple indexes, perhaps due to the simplicity inherent in them. The tendency to build univariate systems is for example documented in Su et al. (2017) and in AghaKouchak et al. (2015). Studies that document multivariate systems include Tadesse et al. (2010), Tadesse et al. (2014) and Wardlow et al. (2012). These are, basically, the outputs from the same operational approach as applied in different contexts and they stand out in their design that incorporates 11 variables derived from oceanic, environment, climate and satellite data.

Closely related to the popularity of univariate systems is the tendency to model droughts using only variables that monitor one type of drought despite the existence of multiple types of drought: meteorological, hydrological, agricultural and socio-economic (UNOOSA, 2015).

Therefore, we observe that with the proliferation of data across the multiple types of drought, future drought monitoring systems should be multivariate and cover multiple if not all the types of droughts.

2.13.3 Single technique- single model in the prediction of droughts

The state of the art in drought prediction is replete with the use of a single prediction technique in the search for the champion model. The champion model is the model judged to be most predictive of a phenomenon under predictive investigation.

Majority of drought prediction studies take the approach to modelling that realizes a single champion model. The champion model approach has been shown to:

- have low predictive power and to be unstable in future performance and thus requires frequent updates.
- happen in a context where there exist multiple drought prediction techniques that have been shown capable of offering good performance in the prediction of future droughts. Drawing synergies from these techniques becomes a natural step in the architecture of future drought prediction systems.
- exist even among studies that make an attempt to deploy multiple techniques but that still ended in the selection of the best technique or the comparison of techniques.

The use of single techniques and the selection of single models even in multivariate systems is a loss of opportunity in model ensembling that synergizes the strengths of diverse modelling techniques to realize both highly predictive and stable models.

2.13.4 Opportunity for model ensembling

Model ensembles have in the past been done using bagging and boosting with the aim to realize good predictive performance from multiple individually poor learners (Belayneh et al., 2016; Opitz & Maclin,1999). While bagging reduces variance by averaging predictions, boosting increases model performance by sequentially learning models. Stacking, which is increasingly becoming the attention of research as an approach to model ensembling aims to combine weak learners to reduce generalization error. Džeroski & Ženko (2004) document stacking as having at worst a performance comparable to champion models.

In theory, model ensembles should outperform single models. Furthermore, heterogeneous ensembles are expected to outperform homogeneous ensembles as documented in Petrakova, Affenzeller & Merkurjeva (2015). This, however, is noted to not hold in the case of some empirical studies like Elish (2013) in which in 2 out 5 instances, single models outperformed model ensembles. In fact, some studies like Kocaguneli, Kultur & Bener (2009) had ensembles lose performance to champion models.

We realize that empirical studies that document the use of model ensembling to realize more predictive models, have themselves not settled on the question of the performance of model ensembles as compared to best performing models. The investigation of the performance of different ensemble approaches, especially between homogeneous and heterogeneous ensembles is a gap in drought monitoring and indeed in many other applications.

We see that despite the predictive performance of model ensembles being at variance across studies, their application would benefit the prediction of droughts with more predictive and stable models. The extent of this benefit would be further enhanced with the analysis and comparison of the performance of the different ensembles relative to one another and with the champion model as the base for comparison. The evaluation of the performance of the ensembles would offer a solid guide to practitioners on what kind of ensembles to go for to be guaranteed of higher predictive power and stability.

Chapter 3: RESEARCH DESIGN AND METHODOLOGY

In this chapter, we outline the study area, the research design, the research process and offer justifications for choices of both the research design and research process. We provide justifications on the question of why we did not choose alternative and closely related research designs. The section, besides, discusses the type of data used for the study, how the data is collected and how nuggets are realized from the analysis of the data.

3.1 Study Area

The study area is shown in Figure 3.1. The study area comprises four counties of Kenya: Turkana, Marsabit, Mandera and Wajir. The selected region lies in the northern part of Kenya that is characterized as part of the arid and semi-arid lands (ASALs) of Kenya. The extent is bounded by: Upper Left X (Lon) 33.918, Upper Left Y (Lat) 5.513, Lower Right X (Lon) 41.967 and Lower Right Y (Lat) 0.147. All four selected counties are classified as arid. The study area is, therefore, part of the ASALs monitored by the National Drought Management Authority (NDMA) of Kenya.

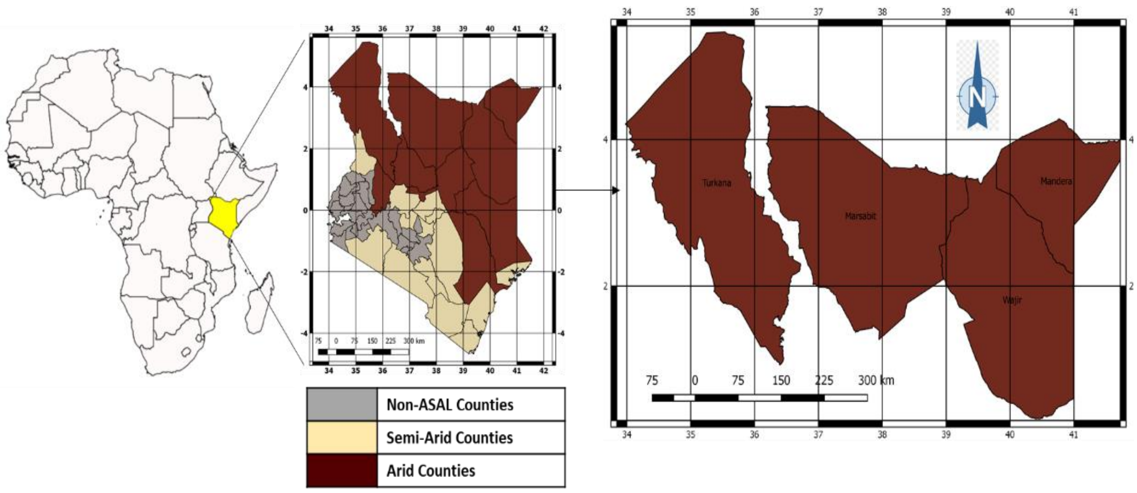


Figure 3.1: Map of the study area

The map shows the location of Kenya in Africa and the counties of Turkana, Marsabit, Mandera and Wajir in Kenya with the map of Kenya and whilst showing the grouping of Kenya counties into arid, semi-arid and non-ASALs.

The counties in this study area cover a combined area of 215,242 km^2 with a total population of around 3.04M (KNBS, 2019). The annual average rainfall for the counties is 250mm for the three counties with Wajir having an average of around 370mm. The rainfall pattern is bimodal with the long rains in March, April and May (MAM) and the short rainy season between October, November and December (OND) with 6 months considered wet months. The monthly average vegetation cover (2003-2015) as quantified by the Normalized Difference Vegetation Index (NDVI) from the BOKU system for operational drought monitoring (Klisch & Atzberger, 2016) is as shown in Figure 3.2.

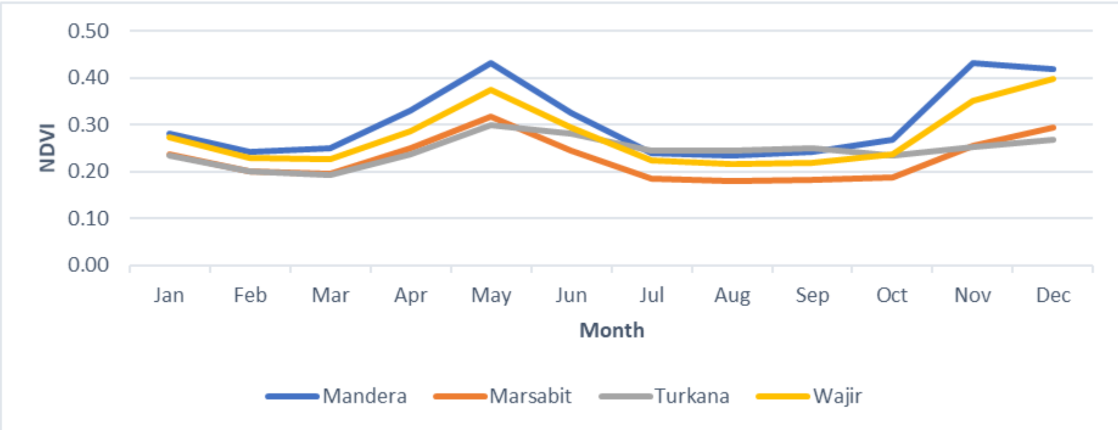


Figure 3.2: Average NDVI across months for the counties in the study area based on MODIS data for the period 2001-2017.

The NDVI data for the period 2001-2017 (Figure 3.2) indicates very low absolute vegetation cover of 0.43 in a scale of -1 to +1 implying very sparse vegetation cover and hence low biomass even during the wettest months e.g. May and December across the counties in the study area.

In addition to the continued occurrence of droughts in the study area, further justification for the choice of the study area are based on the: -

i. Index-Based Livestock insurance product development for the four counties of Turkana, Mandera, Wajir and Marsabit. The index is based on NDVI and will benefit from the study that will offer an alternative source of drought monitoring products that could complement the exclusive use of vegetation data, the NDVI (Chantarat et al., 2013).

ii. The scalability of the Hunger Safety Net Programme (HSNP) that is run in the four counties in the study area is based on drought status as indicated by the VCI3M which is an NDVI based product (Beesley, 2011). HSNP will be a potential beneficiary from this study since the research already developed the scalability model for the programme and a predictive early warning system will be a bonus to the process for scalability in social protection especially if the objective is early preparation in the delivery of the drought shock responsive payments.

3.2 Research Design

As documented in Thompson et al. (2005), all analyses are correlational and all quantitative studies yield correlational evidence. The differences between evidence should, therefore, as advocated in Thompson et al. (2005), be rooted not on how the evidence is analyzed but on the design of the study yielding the evidence. The function of the research design is, therefore, to ensure that the evidence obtained from any given research unambiguously answers the intended questions. The intended research design is therefore meant to identify the key pieces of evidence required to achieve the research objectives. Research design, by the above principles, is therefore superior to the question of data collection methods and requisite data processing strategies.

In this study, the intention was to investigate the relationship between meteorological, hydrological, agricultural and socio-economic drought variables to both drought severity and malnutrition status of children under 5 years as impacted by drought. The predictor variables were used to build two sets of models. First, is the set of pure method models that have one predictive modelling technique while the second is an ensemble of different modelling techniques in the prediction of the two target variables. The predictor variables are an integration of both remote sensing and socio-economic variables. An investigation of existing theories informed the design of tools used for the investigation of relationships between the variables.

Once the tools were developed in the form of scripts, the study deployed them to conduct experiments to quantitatively measure the relationships between different variables and to build and determine the performance of ensemble models in drought monitoring.

With consideration to the above, the study, therefore, fits the logical order of the enquiry in **correlational research design** for the investigation of the relationship amongst the various quantifiable variables and the building of the ensemble drought monitoring models.

3.2.1 Correlational Research Design

Correlational research design is operationally defined as the analysis of co-variant data to determine pre-existing relationships with no attempt to manipulate the independent variable. This research technique is noted in Creswell (2013) and further in Curtis, Comiskey & Dempsey (2016) to be used to relate two or more variables to allow for prediction of an outcome based on the correlational relationships between the variables. In essence, the focus of a correlational study is to find already existing relationships between variables without manipulating the extraneous variables. It is observed variously and particularly in Creswell & Creswell (2017) that quantitative research has two major distinctions. The first distinction in quantitative research is based on the existence of an independent and a dependent variable while the second is the methods used for the research that are either experimental or correlational. Correlational research, given that it a form of quantitative research stands at par, in value, to experimental research.

This study, at the core, was interested in determining variables that correlate with drought severity and with nutritional impacts of drought without necessarily accounting for any external variables that might affect the relationships between the variables. Causation was not intended to hold in the relationship between the study variables given that the predictor variables were in no way manipulated.

Despite correlations between the variables being established while causality is not implied, correlational research design still falls in the domain of quantitative methods of research in which two (2) or more quantitative variables from the same group of subjects are investigated to determine if there is a relationship or covariation between them. Correlational research designs are either explanatory design or predictive design in nature.

- ***Explanatory Research Design (ERD)*** looks for simple associations between variables and investigates the extent to which the variables are related. ERD involves the determination of the correlation between two or more variables and uses data collected from observations of a single group to draw conclusions from statistics alone.
- ***Prediction Research Design (PRD)*** is driven by the goal of realizing the prediction of some variable based on other variables. Variables in this design play different roles: as either predictor variables hence used to make a prediction or as the predicted variable that are the target variables for the prediction.

For this study, the research design of choice is thus characterized as a ***correlational design***. This since we investigate the relationships between variables using explanatory research design and also produce models that are used to predict other variables i.e. drought severity and drought effects on malnutrition in a process that falls within the ambit of prediction research design.

3.2.2 Why is the research design correlational?

Broadly, science has any of the three goals of describing, explaining or predicting. These general goals of research are what are to be achieved by different research designs. The correlational design is one such design that achieves two of these three research goals of describing the relationship between variables especially for explanatory research design, and at the same time supporting the prediction of a phenomenon using other variables. The choice of correlational design for this study is justified by the following: -

- First, a view of this study into two stages realizes the investigation of the relationship between the variables used in the study. The relationships are both between predictor variables (X-X) and between the predictor variables and the predicted variables (X-Y). The study determines the correlation between the remote sensing products of both precipitation and vegetation on one hand and between these remote sensing products and socio-economic data on the other hand. These variables were then used to build predictive models for drought monitoring

- Second, a correlational research design is the measurement of two or more factors to determine or estimate the extent to which the values for the factors are related or change in an identifiable pattern. The measurement of the magnitude of the correlations between the variables was a key undertaking of the study.
- Third, the correlational design provides models for both explanatory research and predictive research design models that fell in the desired analysis path for this study.

The above three observations directly led to our decision for the use of the correlational research design. Both the explanatory research and prediction research models of correlational research design were used in this study.

3.2.3 Limitations of Correlational Research Design & Mitigation

The limitations of correlational design are well documented in McLeod (2018), Mitchell (1985) and Thompson et al. (2005). These limitations include the inability to infer causality to the existence of spurious correlations and difficulty in interpretation due to non-existence of standards. These limitations are summarized as follows: -

- ***The inability to infer causality*** is the greatest limitation of correlational design. It is the most experienced limitation since it comes with the formulation of the correlation design problem. Correlational design cannot infer causality since all they indicate is co-occurrence. The method does not control for differences and therefore has no treatments in addition to not having control groups. This limitation, however, does not affect our study since we are not interested in causality but in the relationships between the different target variables and the predictor variables as realized from a modelling approach.

- ***The existence of outliers*** can influence the direction and strength of observed correlations and sample restrictions of the range that is not reflected in the general population. This then would lead to incorrect imputations out of data-driven studies. The study uses statistical techniques to handle outliers. Though the simplest solution is to eliminate outliers, the possible influence of outliers on correlations is mitigated on through transformations that rescale the values as the method of choice. The normalized values then have the influence of outliers mitigated on.
- ***Sampling limitations*** can result in samples that are not representative of characteristics as those present in the general population. This limitation, however, cuts across may research designs. There is the need to ensure that sampling at any point of the study is driven by and tested for representatives of the chosen samples. The study does random sampling while the modelling approaches, in addition to random sampling used bagging techniques to overcome this limitation.
- ***Assumption of linearity*** is common in most correlational design approaches. The basis of this assumption is that there exist linear relationships between the variables under investigation. Since there exist methods that do not assume linearity and or normality, there is always advisable to test data for linearity before undertaking any modelling on the data. We did the test for normality on our variables and adopted appropriate methods based on the distribution of the data. All our chosen modelling approaches do not assume linearity but are free of dependence on pre-determined distributions.
- ***The non-existence of a standard interpretation*** characterizes the use of the correlation approach to research. There are no set rules for interpreting correlation coefficients but the use of statistical significance. There is never a standard to quantify the difference in the values of the measures of statistical significance. The background knowledge of research on the data, therefore, is

heavily relied on in the interpretation of the results of correlation studies. In the study's modelling approach, we used the determinant of correlation(R^2) and the associated rules of thumb to determine the strength of noted relationships.

- ***Spurious correlations and the directionality problem*** are also significant limitations of the correlational design. Spurious correlations do exist in most datasets. These are correlations of variables that in reality have no relationships between themselves. The spurious correlations are thus correlations by chance. They lead to the inferences of relationships that exist in the modelled world but not in the real world. In some cases, even the fact that one variable affects another does not exactly come with the guarantee on the direction of the relationship. This study is susceptible to spurious correlations given that it is entirely based on time series data. To limit susceptibility to the problem of spurious correlations, the study uses two approaches. First, we have surveyed literature and all our variables have a past basis of use in the prediction of drought. Second is the de-trending of the time series data before comparison to reduce the chances of working with spurious correlations.

3.2.4 Positives of the Correlational Research Design

The positives of the correlational research design can be viewed in the context of the advantages of observational research, as compared to other research designs. The positives of correlational research design include prediction, navigation on ethical issues, support for multivariate analysis and ability to work for non-experimental designs.

- *Predicting one variable from another* is a positive offered by the correlational research design. This is the case since the correlational research design provides a framework for the prediction of one variable from another with some considerable accuracy.
- *Overcoming ethical considerations* in the study of sensitive phenomena can be done via correlational design as opposed to, say, actual experimental research design. This is since identities can be reduced or reduced to data in non-

intrusive ways without the creation of groups and even the use of human subjects. Since the results are grouped, there is no direct reference on individual elements and thus privacy violations rarely exist. This study benefits from this in two aspects. First, the predictor variables cannot be manipulated. Second, if they were, it would have been unethical to create drought so as to be able to register changes on it as a result of changes in the predictor variables.

- *Multivariate variables support* ensures that multiple variables can be used in the prediction of one predicted variable. Multivariate systems then guarantee the realization of complex relationships between datasets and data items.
- *The support for non-experimental studies,* especially for some phenomena that cannot be studied by experimentation makes the correlational design a powerful research design. It is, however, this simplicity of approach that leads to the limitation to infer causality.

3.2.5 Why not the closely related Experimental Research Design

In this section, we justify why the alternative and closely related experimental research design was not made a method of choice. True experimental research design as documented in Creswell & Creswell (2017) is the only research design that can establish a cause-effect relationship within a group or groups. It is hallmarked by three factors that need to be satisfied:

- The presence of a control group that is within the research but do not have the research experimental rules applied to them.
- The existence of a variable that can be manipulated by the researcher
- The element of randomness in the distribution of participants between the experimental and the control groups.

The justification for use of the correlational design over the experimental design is thus discussed under the key differences between experimental and correlational research designs based on the: manipulation of variables, types of inferences and the presence or absence of groups or levels of analysis.

- *Manipulation of the independent variable(s)* is the hallmark of experimental research. The independent variable is manipulated to produce one or more results called dependent variables. This is as opposed to observation-based research designs like correlational research designs that do not manipulate any variables and thus refer to the variables as either predicted or predictor variables. In this study, since we use historical data that was strictly observational and without any manipulations of the predictor variables, the correlational research design was the logical option.
- *Holding variables constant* in experimental research ensures that a variable is controlled-for so that by looking at the effect of one, all others are assumed constant in the experimentation process. This is in direct contrast to the correlational research design that does not hold any variables constant in the investigation process.
- *Causal Inference* is possible in experimental research design as opposed to correlational research design. Experimental research has both dependent variables (DVs) and independent variables (IVs) and accounts for external variables (EVs). Cause-effect inference can, therefore, be made from the investigated relationships.
- *Groups/ levels* are used in experimental research design since it has both treated and untreated groups in the research process. The design also deals with the question of random assignment and the counter-balancing of groups. This is unlike the correlational design that just concerns two or more variables.

The above differences, given the fact that we use archival data that were a result of observation processes, account for the study being justified in choosing correlational research design over experimental research design. The study does not manipulate the independent variables, does not hold any variables constant, does not infer causality and has no grouping for the variable under investigation for relationships.

3.2.6 Assurance of validity in correlational research design

The validity of research is either internal or external and deals with the question of the soundness of research as influenced by both the research design and research methods. The question of validity and reliability is for example dealt with in Mohajan (2017) and earlier in Mitchell (1985).

Correlational research design is documented to be poor at internal validity due to the absence of variable manipulation. This limitation is amplified in the case of this study since, specifically in our case, the design of data collection is not controlled since observation data from existing data repositories are used. This absence of design of data collection instruments and procedures then is a threat to internal validity. Despite this limitation, the soundness of data collected is one that was done as part of this study to ensure good quality data is used and that data manipulation processes were sound.

Given that the internal versus external validity question is a trade-off and that external validity concerns the question of generalizability beyond the study group, it is the case that correlational studies have higher external validity as compared to experimental research design. This is attributed to the fact that correlational studies do not control for variables and so typify reality as compared to other approaches. Despite this assurance, we ensure populations are chosen randomly to avoid sampling bias and we also use a test data-set that is from the future to test for continued validity in the face of generalization.

3.3 Description of the Research Method

While Research design defines the blue-print for a study based on the steps to be taken, Research methods refer to the techniques deployed in research to gather information. In the assertion of Macdonald & Headlam (2008), research methods are either quantitative or qualitative and are defined to involve the quantification of things and ask the questions of how much, how long and to what extent. Qualitative research, on the other hand, concerns the quality of information. The data in quantitative research can thus be counted, sorted, measured, classified etc. The research method is discussed in terms of the data collection and data pre-processing and the chosen methods of data analysis. The quantitative research process and the correlational research design were together operationalized in this study following on the research process discussed here-next.

3.4 Research Process

The Research Process documents the process as defined by the set of finite steps we followed in undertaking the research study. With the research design already set as correlational research design, the basic steps of this research followed on this design. We present the steps and how they fit in this study under the design of investigation (DoI), logical flow of the process and finally provide a model for the study process.

3.4.1 Design of Investigation

The study uses both explanatory (descriptive) and predictive aspects of correlational research design. The investigation of relationships culminates into the effort to build predictive models that try to approximate the target (criterion) variables using the predictor variables. The steps to be followed based on the research design are as shown in Figure 3.3.

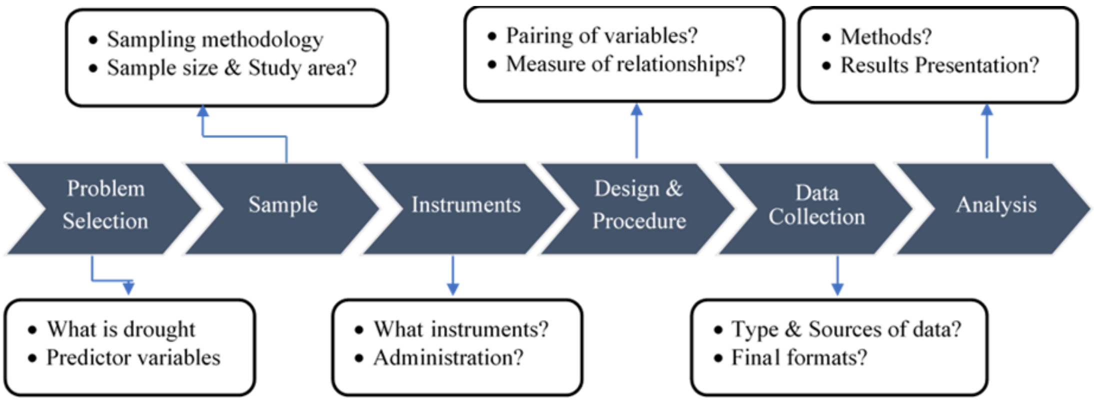


Figure 3.3: The correlation research process.

The different components of the correlational research process (Figure 3.3) are addressed at different phases of the study from problem formulation to results documentation and interpretation. Each step has the key questions formulated. The steps in the research process are as described below:

- ***Selection and definition of the problem*****:** The selection of the problem and its definition is the first undertaking of the correlational research design. We have defined the problem as the identification of the remote sensing and socio-economic variables that are used in the monitoring/ prediction of drought and to investigate their relationship with both drought severity and drought effects. The identified variables are then used to build ensemble models of both ANN and SVR whose performances are then compared to the traditional champion models. This was as defined in section 1.4 and section 1.5 and is formalized in terms of methodology in section 3.5.1 on Methodical definition of the problem.
- ***Sample:*** The question on the sample size and sampling methodology is dealt with under the section 3.6.3 on modelling methodology while the question of the study area was dealt with in section 3.1 as a purposive selection that was based on the expected utility of the outputs of the research. The key issues in the definition of samples and sizes are discussed as part of the modelling methodology.
- ***Instruments*****:** The study uses data extracted from data repositories and operational drought monitoring systems to achieve its objectives. The instrumentation, therefore, can be equated to the development of scripts and program tools that are used in the extraction of required data elements that are then pre-processed to realise the variable/ indicators necessary for the study. Instruments are discussed together with data collection in section 3.5 on data collection.
- ***Design and procedure:*** The design of the correlational design, in this study, is considered a straight forward task. In the exploratory phase, scores or observations between two variables are paired and their correlation coefficients calculated to indicate the degree and direction of the relationships. In the predictive phase, the aim was to build models and their combinations that best predict the target variables given the prediction variables. The experimentation was set-up to follow the machine learning process as reviewed earlier in Figure 2.10 in section 2.8.1. The section on modelling methodology covers the

elements of the design of experimentation as well as the section on data collection and exploratory analysis.

- ***Data collection:*** The issues encompassing the process of data collection including the data requirements, the data sources, data types, native/source formats and format conversion to the requisite final formats are handled in the section on data collection. The process for handling missing data and ensuring validity is also presented
- ***Data analysis, presentation and interpretation:*** The statistical inferences of the derived relationships and predictive models and their associated interpretations are presented under the chapter on results and discussion to draw key answers to the objectives of the study.

3.4.2 Logical flow of the study

To achieve the objectives of the study, the logical flow of the study is expected to follow on the steps as presented in Figure 3.4. Through presented as a linear process, in implementation the process is expected to be massively iterative until the questions are answered and objectives all met. The different components are handled in different phases of this research.

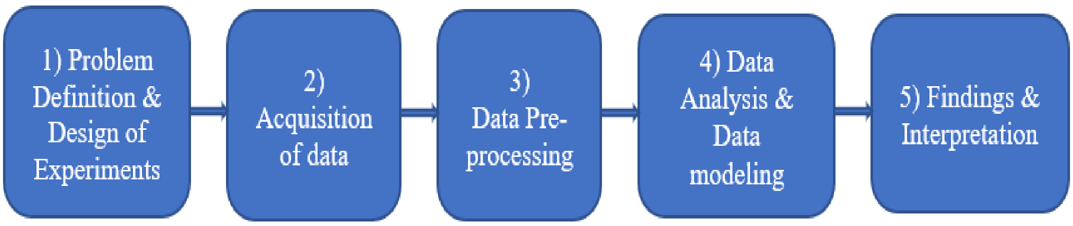


Figure 3.4: Logical flow of the study.

3.5 Data Collection and Data Pre-processing

3.5.1 Overview of Data Acquisition

Data mining approaches essentially are best suited to vast volumes of data from which non-trivial information is sought. The common approaches to the acquisition of data for data mining are either direct observation over a period or the use of archival data.

Archival data is extracted off from some past observations, typically from operational systems or data-warehouses. The study extracted data from multiple data archives. The decisions on data acquisition were majorly driven by the deployed type of research for the study. Based on Miles & Huberman (1994), this research is considered to be quantitative research. This is justified by the fact that we aim to count & classify features and construct statistical models in an environment in which the study design informed data collection and data is in the form of numbers and statistics.

In this study, data collection is widely referred to as *data acquisition*. The process of data acquisition and data pre-processing is non-trivial. The non-triviality of data acquisition is attributed to differences in data quality issues, data format issues and accessibility issues around the different sources. The socio-economic and remote sensing data are both treated as time series data collected at a monthly frequency over the period March 2001-2017 for the remote sensing data and at a similar frequency for the socio-economic data over the period 2008-2017 as illustrated in Figure 3.5. While the remote sensing data is from multiple sources, the Socio-Economic data is sourced from both the legacy system and the active operational drought monitoring system of the National Drought Management Authority (NDMA). The socio-economic data is shown to have a shorter historical archive as compared to the remote sensing data.

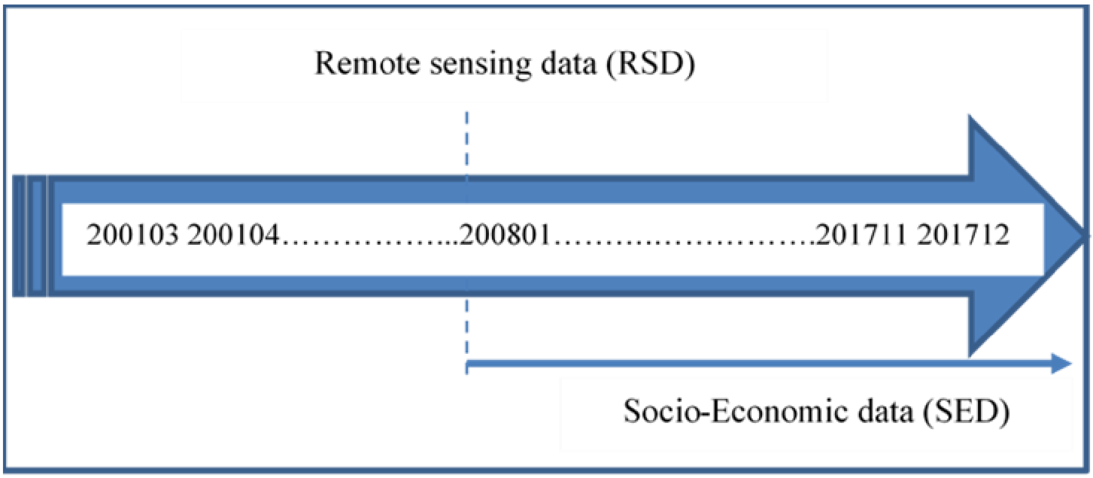


Figure 3.5: Timelines for the major categories of the study datasets.

The extraction of the socio-economic data and the remote sensing data from existing operational data archives, databases and data warehouses all make the research method to also fall within observational research. All the datasets were integrated into one consolidated dataset.

The socio-economic data was transformed so as to create more variables deemed sensitive to drought. Such transformations were aimed at exploring the possibility of ending up with more predictive variables from the initial set of variables.

3.5.2 Key data collection/ acquisition concepts

The data collection process is in cognisance of the following issues: -

- ***Unit of Analysis:*** The unit of analysis is defined as the county level for the study. The choice of the county as the unit of analysis is justified by two key considerations. The study accommodates both socio-economic and remote sensing data. The socio-economic data is collected, analysed and reported on at different units of community, livelihood, sub-county and county level. This coupled with the fact that droughts affect relatively large areas justifies the choice of counties as the unit of analysis to avoid excessively rapid changes that would not be convenient for drought monitoring.

  Though remote sensing data could be processed at smaller spatial units than counties, the choice of the county level was, therefore, to ensure that there exists uniformity in spatial coverage between the remote sensing data and the socio-economic data.
- ***Ethical Considerations:*** According to Slonecker, Shaw & Lillesand (1998) and factoring in the recent innovations, it is clear that ethical issues abound in the acquisition and use of remote sensing data. Equally, the use of socio-economic data collected at household levels is subject to possible ethical violations. In this study, we use remote sensing data that is captured at 250m resolution. The data is supplied as images with digital values that do not have any possible violations to privacy or any other such infringements.

  For the socio-economic data, we sourced the data from the existing archives. The field-based data collection process goes through the listing of households and the seeking of consent from household heads for monthly data collection. Household opt-out of data collection process unhindered and thus participation remains voluntary and with consent. Moreover, for this study, we use and combine both the approaches of depersonalization of the data to ensure it is without any identifiers and also use summarized statistics at the county level. This makes the data non-traceable and non- constructible back to the original interviewees.

- ***Guiding Principles used of data collection:*** The process for data acquisition in this study was guided by the need to acquire relevant data that adds to the intended body of knowledge, reliability of the said data and data that remains valid and thus able to measure what is intended to be assessed. The data acquisition systems and processes were made systematic and well documented for reliability and reproducibility. The major questions that were tabulated for each data source and evaluated before actual data acquisition were as outlined in Table 4.

Table 4: Data sources evaluation criteria

Consideration	Evaluation
Accessibility of identified data	All data sources identified as potentially accessible were subject of data collection
How to finance data collection	Some data collection processes required investment in the actual data acquisition. These were evaluated in the context of cost versus alternative sources
Number of potential variables	We evaluated all possible variables and their transformations to determine their suitability for the modelling process
Kind of data	We ranked Primary data sources higher over Secondary data sources wherever both existed.

Since the data collected is categorized into either remote sensing or socio-economic data, we present a description of the data following on these types. For remote sensing data, we adopt their discussion based on the types of drought as provided earlier in Figure 2.5 and adopted study conceptual framework in Figure 2.20. Similarly, we discuss the socio-economic data following on their expected lag to drought.

3.5.3 Remote Sensing Data Collection

Two things are important to note about the remote sensing images that are used in this study: -

- ***First*** is that the remote sensing data used in the study came in two distinct formats (HDFS) for all the datasets except for SPEI that comes in NetCDF format.
- ***Second*** is that the images have multiple bands. The processing of remote sensing indicators from the images demands the use of specific bands of the data.

Table 5 provides a list of the bands, if available, in the different images/ remote sensing datasets and their corresponding bands of interest. The process for data collection for remote sensing data in this study was equated to the download of the data from identified data repositories. The remote sensing data acquisition process was achieved by identifying the data, the data sources, initial data formats, the bands in the data and the bands of interest thereof. In the implementation, the download process is automated for the identified images using R scripts.

Table 5: Remote sensing dataset sources, bands and description

Dataset Name	Dataset Source	Description	Science Dataset (SDS) Layers/ No of Bands	Spatial Resolution (M)	Temporal resolution (days)	Bands of Interest	Description of Bands
MOD13Q1 & MYD13Q1	Didan (2015a) & Didan (2015b) respectively	MODIS-Terra and Aqua vegetation indices respectively	12/36	250	16	3	16-day NDVI average VI quality indicators Day of year VI pixel
MOD11A2	Wan, Hook & Hulley (2015)	MODIS- Terra Land Surface Temperature/Emissivity	12/36	1,000	8	4	Daytime LST & Quality Nighttime LST & Quality
MOD16A2	Running, Mu & Zhao (2017)	Terra MODIS Evapotranspiration	5/36	500	8	2	ET & PET
SPEI	Beguería et al (2014)	Standardised Precipitation-Evapotranspiration Index	-	~55,500	30/1	1	SPEI
TAMSAT	Tarnavsky et al. (2014) Maidment et al. (2014) and Maidment et al. (2017)	TAMSAT Monthly rainfall estimate	-	4,000	30/1	1	TAMSAT RFE
CHIRPS	Funk et al. (2015)	CHIRPS monthly rainfall estimates	-	5,500	30/1	1	CHIRPS RFE

3.5.3.1 Remote sensing data pre-processing

The study aims to realize a time series data for each of the four counties in the study area: Mandera, Marsabit, Turkana and Wajir. The boundaries of the counties as administration units are defined by the Kenya counties shapefile source from the Independent Electoral and Boundaries Commission of Kenya. The intent was to have the statistics at a monthly frequency. The generic steps in the pre-processing of the remote sensing data are as shown in Figure 3.6.

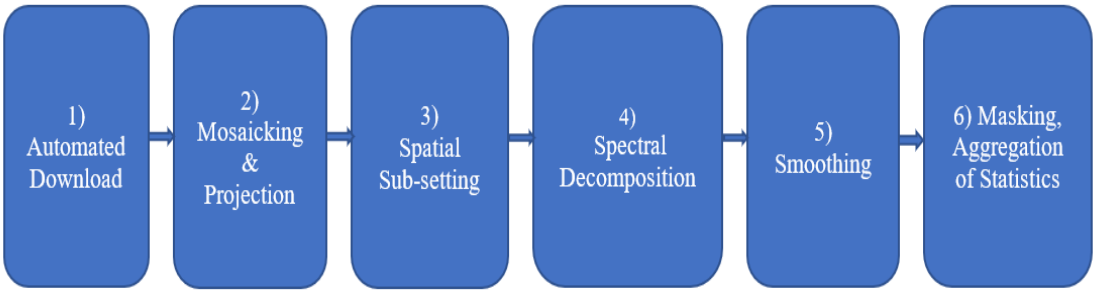


Figure 3.6: Steps in the pre-processing of remote sensing images.

The study methodology for the processing of remote sensing data follows from literature, and though with differences, closely monitors that documented in Klisch & Atzberger (2016). The pre-processing steps as implemented are described below. The pre-processing steps were all done using scripts written in R.

- ***Automated data download and spectral sub-setting***

  The data downloaded for all the remote sensing data straddled the period March 2001 to December 2017 but at different frequencies for each dataset. The download script involves iterating for all the appropriate images for each dataset within the study period. Due to storage and download speed concerns in **Step 1** and given than most of the download was mostly for multi-band images, the study downloaded only a sub-set of the bands. The downloads process, therefore, incorporated an element of spectral sub-setting similar to that in **Step 4** even though the resultant image were still multi-band. The images were in either of HDF format (MOD13Q1, MYD13Q1, MOD11A2 and MOD16A2) or netCDF (SPEI, CHIRPS and TAMSAT). The MODIS images were sourced from the data archives: Land Processes Distributed Active Archive Center (LP DAAC, 2014) and Level-1 and Atmosphere Archive & Distribution System Distributed Active Archive Center (LAADS DAAC, 2016). The download script is based on the R MODIS Package while the re-projection is done using the MODIS reprojection tool (MRT) documented in LP DAAC (2004). A sample of the settings for the automated download of scrips has the parameters presented in Table 6.

Table 6: MODIS Terra Download Parameters

Characteristic	Description
Product Code	MOD13Q1
Collection/ Version	006
Extent	h=c (21,22), v=c (8,9)
Spatial resolution	250M (x and y)
Science Data Set String	100000000011
Temporal resolution	Begin=”200103”, End=”201712”

The sample indicates the parameters to download version 6 of the MODIS Terra data for the period of study for the four tiles covering the Kenyan extent. This process produces one tri-band image for Kenya with the three layers of choice: 16 days NDVI average, composite day and pixel reliability. Potentially, the spectral subset image is reduced in size to almost 9% of the original size. This leads to benefits in download time, subsequent pre-processing time and storage requirements. Spectral sub-setting at this stage is thus an initial *data reduction strategy.* The tiles downloaded to cover the entire Kenyan extent are as shown in Figure 3.7.

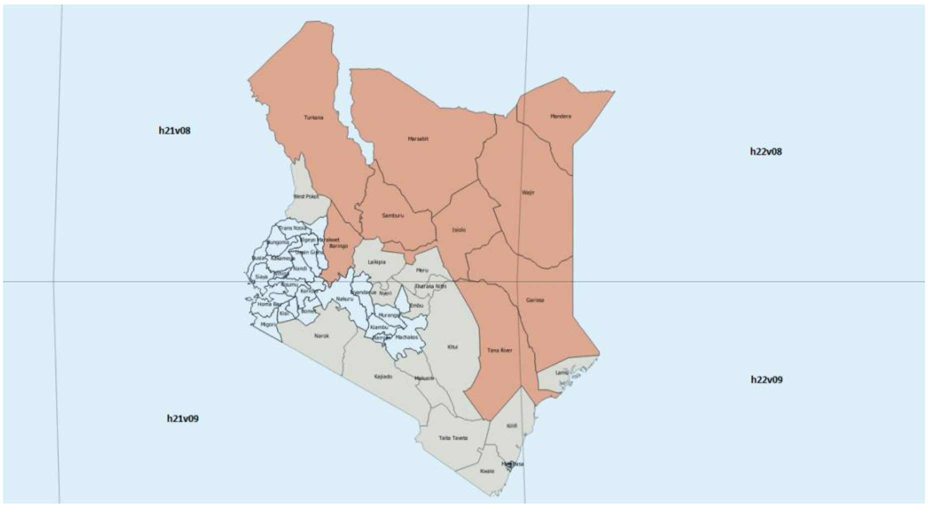


Figure 3.7: Download MODIS data tiles covering the entire Kenyan spatial extent indicating the relevant tiles - H21V08, H21V09, H22V08 and H22V09

- ***Mosaicking, Re-projection and Re-sampling***

For some of the data, like the vegetation datasets, multiple tiles are downloaded for each spatial extent of data download. The case of MODIS data including vegetation data (VGT), evapotranspiration (EVT) and Land surface temperature (LST) comes in 4 tiles for the Kenyan extent for each time step of download. These tiles are then stitched to form a single image for each time-step in a process referred to as mosaicking. The study downloaded the data for the entire Kenyan extent since the remote sensing process preceded the socio-economic data process and therefore there was the need to forestall the effects of possible changes based on the availability of appropriate socio-economic data. The single value decomposition (SVD) approach was followed for the decomposition of the images in the study. Given the possible effects of clouds as reducing the digital values of an image, maximum value decomposition (MVD) was chosen as the method of composition of the images to form the mosaicked tiles as illustrated in Figure 3.8.

The study re-projected the images from their *projection* (*coordinate reference system (CRS))* to the geographical coordinate system that is the chosen coordinate system for the output images. The MODIS based datasets, for example, come in the sinusoidal projection that is then converted to geographic coordinates based on location on a plane (longitude, latitude). We use the nearest neighbour (ngb) resampling method as implemented in R in the interpolation of cell values to the projected raster images.

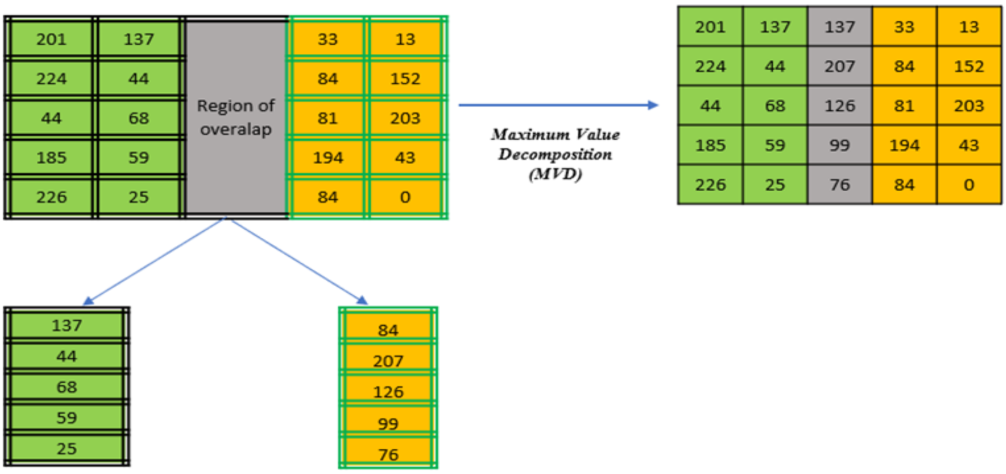
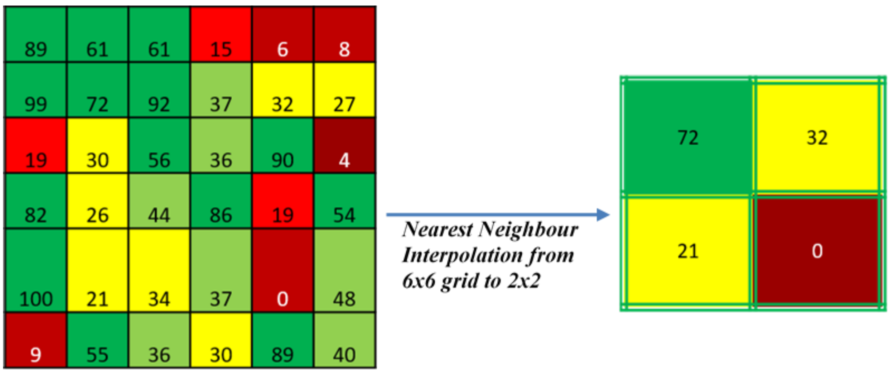
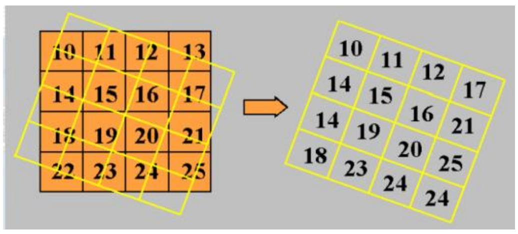


Figure 3.8: An illustration of mosaicking using maximum value decomposition (MVD)

Nearest neighbour resampling was used on only the pixel level data on the regions of overlap to determine the values to the output cells based on the nearest cell centre to the input grid as illustrated in Figure 3.9.



(a)



(b)

Figure 3.9: Nearest neighbour method of interpolation.

The Figure 3.9 illustrates (a) the use of nearest neighbour in the resampling of remote sensing images with the output cells centred for each set of 3 cells square and (b) resampling where re-projection has occurred and therefore cell values change. In either case, the final cell values are all from the initial image domain.

- ***Spatial sub-setting***

  The images underwent spatial sub-setting that entailed the reduction of the spatial extent to the area of interest that is less than the rectangular coordinates of the downloaded images. The reduction was to the spatial extent of the study area using the extent derived from the rasterized shapefiles of the study area similar to that provide earlier in Figure 3.7. Spatial sub-setting served as a further data reduction strategy that makes for a reduced dataset to smoothen and query as undertaken in the next set of remote sensing data pre-processing steps.

- ***Spectral decomposition and smoothing***

  The NDVI images having been spatially sub-set to the required spatial extent of the study area were then spectrally sub-set into the three layers: 16-day average NDVI, VI usefulness and composite day of the year. The approach of the study involved using the MODIS package in R to decompose the data to daily images that we then smoothed using the naïve implementation of Whittaker smoothing as implemented in R. Each of the pixels was weighted based on a process similar to that in Klisch & Atzberger (2016). The weights used were based on the VI usefulness values as shown in Table 7. The naïve attempt is in our opinion not capable of performing as compared to the entrepreneurial option used by BOKU (Klisch & Atzberger, 2016) in their operational monitoring. The effect of smoothing was to penalise NDVI values for pixels with higher cloud cover as compared to those captured with better clarity. This is an important process especially to ensure better and reasonableness of aggregated data and avoid the tendency to record low NDVI values solely attributable to cloud cover. The pixels decomposed with cloudy conditions had their NDVI values penalized as opposed to those of less cloudy conditions.

Table 7: Translation of MODIS VI usefulness quality flags to weights for smoothing

MODIS QF	Weight in Filtering
0,1,2,3	1
4	0.8
5	0.6
6	0.4
7	0.2
8-15	0

The implementation in R uses the algorithm provided in Appendix A that is used to convert the VI usefulness image to a weight image used to weight the NDVI image based on the quality of observations as influenced by cloud cover. The implementation was dynamic with the range of linear scaling provided as arguments to the function.

- ***Extraction and aggregation of statistics***

The final step in the pre-processing of remote sensing data is the process of extraction of statistics for the regions of interest and the subsequent aggregation of the same statistics to generate the desired indicators at a monthly level.

All the images after smoothing give an image as their output for every time step of processing. A month could have multiple images depending on the dataset. The process of creation of the single image corresponding to each month is undertaken in two distinct approaches.

Non-transformed data: The first approach to aggregation is for the collation of non-transformed indicators. The non-transformed indicators include RFE, NDVIDekad, LST, EVT, PET. For this category, all the images within each month are aggerated based on the most logical aggregation function on all the pixels to generate a single image. The functions include summation for rainfall estimates, averaging for vegetation like NDVIDekad, averaging of both day and night temperature to daily temperature and subsequent averaging to monthly mean temperature. Averaging to mean monthly values is also done for EVT and PET.

Transformed data: The second approach is for the transformed datasets. The periods of transformations done are either for one-month in temporal coverage or for three months. The transformations are standardised values using either the standardised difference approach or the relative difference approach. We use the approach similar to the approach in Equations 7 and 8 in Table 2. The transformations are however re-defined based on the data as follows:

- Mean and standard deviation i.e. MEAN(P) and StDEV(P) in the standardised difference approach represent the mean and standard deviation month-on-month for the historical period set as 2001-2013 for the datasets.
- Minimum and maximum i.e. MIN(P) and MAX(P) in the relative difference approach imply the minimum and maximum month-on-month for the historical period set for 2001-2013 in the datasets.
- The differencing is at the pixel level and precedes aggregation.

Finally, the process for aggregation is done using the rasterized image for the area of study that is used to demarcate the spatial extents of aggregation. The administrative units' image is resampled to have the same spatial resolution and extent as that of the images by having all the non-covered matching cells set to NA values before summarization by admin unit is done as illustrated in the Figure 3.10.

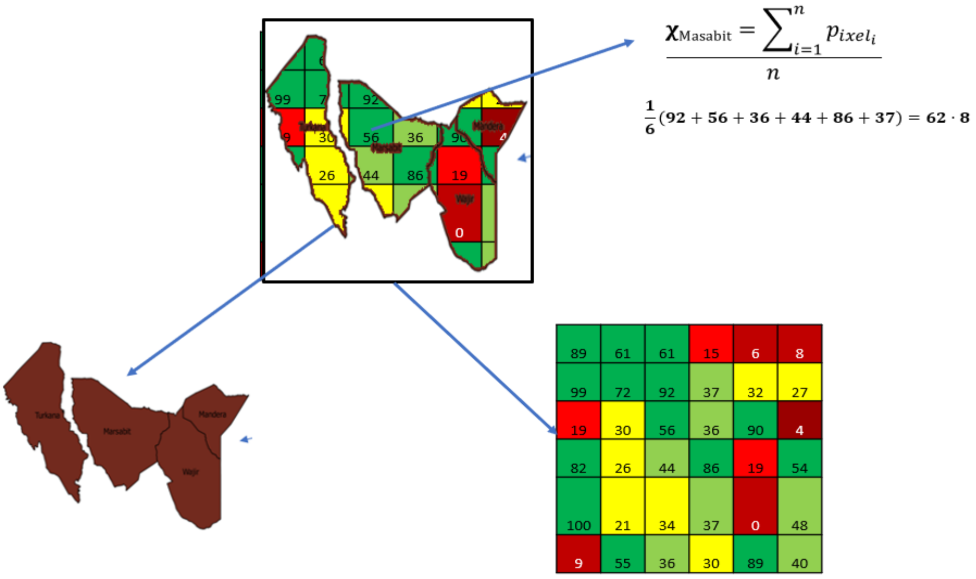


Figure 3.10: Aggregation of pixel data by administrative unit using the average function

The end process of remote sensing data pre-processing entailed logging the statistics in a CSV file format. For each admin unit, there is an entry for each month in the study period for each statistic calculated to form the time series for the remote sensing data.

3.5.3.2 Presentation of Remote sensing variables

The final output of the remote sensing data is in the form of a time series data with the structure shown in Table 8. The information on the types of lagged variables is provided for each of the remote sensing variables.

Table 8: Description of the remote sensing variables

Variable	Variable Description
VCI3M	*VCI aggregated over the last 3 months*
NDVIDekad	*NDVI for the last dekad of the month*
VCI1M	*VCI aggregated over the month*
VCIdekad	*VCI for the last dekad of the month*
TAMSAT_RFE1M	*TAMSAT Rainfall Estimate aggregated over the month*
TAMSAT_RFE3M	*TAMSAT Rainfall Estimate aggregated over the last 3 months*
TAMSAT_RCI1M	*TAMSAT Rainfall Condition Index aggregated over the last 3 months*
TAMSAT_RCI3M	*TAMSAT Rainfall Condition Index aggregated over the last 3 months*
TAMSAT_SPI1M	*TAMSAT Standardized Precipitation Index aggregated over the last 1 month*
TAMSAT_SPI3M	*TAMSAT Standardized Precipitation Index aggregated over the last 3 months*
LST1M	*Land Surface Temperature aggregated over the month*
EVT1M	*Evapotranspiration aggregated over the month*
PET1M	*Potential Evapotranspiration aggregated over the month*
TCI1M	*Temperature Condition Index aggregated over the month*
SPEI1M	*Standardized Precipitation Evapotranspiration aggregated over the month*
SPEI3M	*Standardized Precipitation Evapotranspiration aggregated over the last 3 months*
CHIRPS_RFE1M	*CHIRPS Rainfall Estimate aggregated over the month*
CHIRPS_RFE3M	*CHIRPS Rainfall Estimate aggregated over the last 3 months*
CHIRPS_RCI1M	*CHIRPS Rainfall Condition Index aggregated over the last 3 months*
CHIRPS_RCI3M	*CHIRPS Rainfall Condition Index aggregated over the last 3 months*
CHIRPS_SPI1M	*CHIRPS Standardized Precipitation Index aggregated over the last 1 month*
CHIRPS_SPI3M	*CHIRPS Standardized Precipitation Index aggregated over the last 3 months*

3.5.4 Socio-Economic Data Collection

The socio-economic data as collected by the National Drought Management Authority (NDMA) already has county level as the spatial extent of coverage. The smallest unit of data collection remains the household that resides in a community that lies within a livelihood zone. The raw data collection that we use in this study covers the period June 2005 to December 2017 at a monthly temporal resolution. The hierarchy for the collection of socio-economic data collection is provided in Figure 3.11.

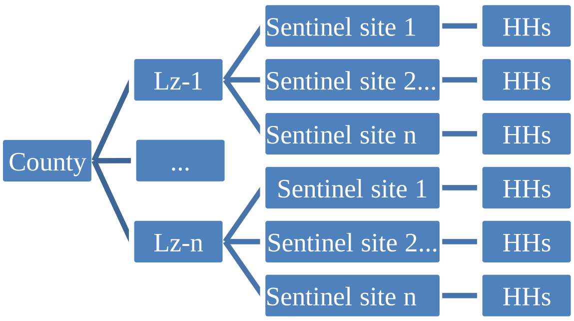


Figure 3.11: The socio-economic data collection hierarchy used by the NDMA. Lz-1 to Lz-n demote the livelihood zones while HHs are the households within the counties in the study area.

The socio-economic data is collected at two points: at the households and at the community level. The community level is, however, a logical structure which identifies the key informants. There is a similarity of some indicators at both the household and the community levels with the community level data acting as controls and validations for such indicators. Table 9 describes the instruments of data collection.

Table 9: Instruments of socio-economic data (SED) collection

Form	Description
Household Administrator (HHA)	The HHA form is administered at the household level and majorly collects data on production, market access and welfare indicators (food consumption and utilization). The form is administered to sampled households consistently every month to the household administrator.
Key Informant Administrator (KIA)	The KIA form collects information on key community-level indicators and is responded to by selected community key informants. The selection of the informants is assumed to ensure people knowledgeable on the community affairs are the key informants. More than one respondent is usually chosen per community and an aggregation done for the responses.

The socio-economic data is collected using two forms: the household administrator (HHA) form and Key Informant Administrator (KIA) form. The KIA is a uniform form in its presentation. The KIA form, therefore, has one record in a de-normalized table for every interviewed Key informant. The HHA form, however, is presented as two different forms: HHA for household-level data and Mid-Upper Arm Circumference (MUAC) form. The relationship between the forms and the respondents is as shown in Figures 3.12a, 3.12b and 3.12c.

From Figure 3.12a, the relationship between the respondents and KIA is one-to-one (1:1) implying one record for every household respondent. The KIA form in Figure 3.12b just like the HHA is also presented as a one to one (1:1) relationship between respondents, key informants, and the filled-in records. The MUAC form, however, has a one-to-many (1:M) relationship between the respondents and household MUAC records (Figure 3.12c). The interview instruction provides for a maximum collection of five MUAC measurements from children under 5 years of age.

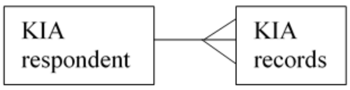


Figure 3.12(a): KIA respondent - KIA records relationship

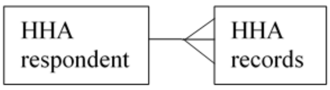


Figure 3.12(b): HHA respondent - HHA records relationships

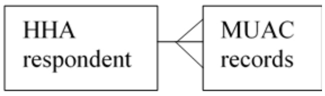


Figure 3.12(c): HHA respondent - MUAC records relationships

3.5.4.1 Socio-Economic data pre-processing

The data pre-processing steps for the socio-economic data is designed to ensure the variables identified are both complete in data and are sensitive to changes in drought conditions. The socio-economic data pre-processing steps are as shown in Figure 3.13.

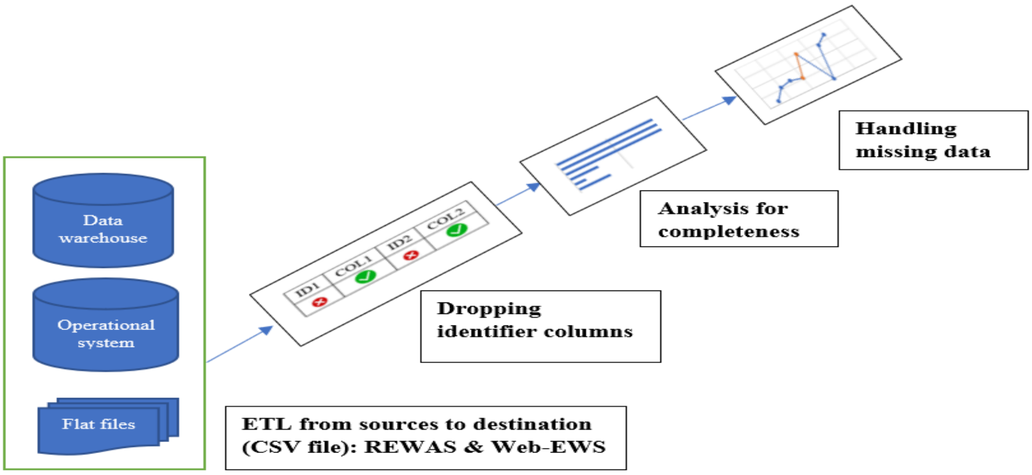


Figure 3.13: The socio-economic data pre-processing steps.

The socio-economic data pre-processing has the steps indicated in Figure 3.13 including: *(1)* data extraction, transformation and loading (ETL), *(2)* elimination of identifiers and non-drought sensitive variables, *(3)* analysis for completeness and gap-filling. Though indicated as a linear process, the last two processes on analysis for completeness and gap filling are particularly iterative till some pre-set goal of completeness is achieved. The methods employed in each of the steps in the process are described as follows:

- ***Extraction, Transformation and Loading from data sources to flat file***

  The extraction of the data involved getting data from the legacy systems of the NDMA, the current operational drought early warning system and a review of drought bulletins and food security assessment reports for the counties on the reports on the major indicators.

  The legacy system of the NDMA, the Revised Early Warning System (REWAS) is a text-based graphical user interface-driven system. The data comes in a comma delimited text file that accumulates the data from across the years of its operations. The data was stored in multiple files in this database for the months covering June 2005 to June 2016. Thee period July 2016 to December 2017 has its data collected from an internally developed web and mobile-based system that was pioneered by the authorship as a result of the analysis of the limitations of the previous REWAS. The new system is a relational database-driven system.

The data from the past and present operational systems were finally integrated into a comma-separated value flat text file as the desired data input to the R scripts that were used for data processing and model building.

- ***Elimination of identifier and non-drought sensitive variables***

  The initial stages of data pre-processing were designed to eliminate variables that had no relationships to drought monitoring as viewed from a drought monitoring perspective. It was thus guided by considerations for possible information value of columns were they to be used as variables. This initial data reduction was done in two stages: removal of identifiers and subsequently dropping those with no expectation of contribution to drought prediction.

  - The initial reduction included the removal of record identifier fields and administration unit identifiers below district level including district, division, location. This was followed by the depersonalization of the data to ensure names and interviewer identifiers are dropped off the data. The process for the three different SED sources saw, for example, the reduction of MUAC variables from initial 22 to 7 columns; HHA variables from 208 to 195 and 195 and KIA from 213 to 196 columns.
  - The second phase of the initial reduction of data volumes was based on the expert knowledge on the data being pre-processed for modelling. The key decisions made in the process included the dropping of variables that seemed not to offer any information based on the target variable which for this study is the change in drought conditions. Variables that were more tuned to the monitoring of food security, rather than drought monitoring, like food stocks and incomes were also dropped since they are not responsive to changes in drought situations. The process of elimination of non-drought responsive variables realized, for example, the reduction of MUAC variables from an initial

7 to 1 with two identifier columns. The HHA variables were reduced from 195 to 24 and KIA from 196 to 15 variables.

- ***Analysis of completeness for socio-economic data***

  The analysis for completeness involved setting the criteria that define the completeness of data. The criteria were set as follows:-

  - An interview must have been conducted in at least 50% of the households from all the sentinel sites covering the county. The number of households was based on the reduced numbers based on the 2017 revision that had 30 households per sentinel site and the number of sentinel sites for the counties was: Mandera (11), Marsabit and Mandera (9) each and Wajir (8) with each sentinel site having 30 households.
  - The sentinels must cover all the livelihood zones of the county. Due to the homogeneity of the livelihood zones in the study area, the condition was that interviews must have been done in at least all the livelihood zones.
  - For all the column values to be filled in, at least 50% of the expected records had to be provided without cases of missing data for representative aggregation.
  - A month was thus considered complete in its data if and only if all the three conditions above we met.

With the conditions above, the years June 2005 to June 2008 were vastly considered to have lots of broken data collection processes and thus the data for the period were not used in the study. The major cause of data issues in this period is down to data loss as a result of the loss of database files of the legacy system REWAS together with the difference on go-live times and the existence of test data within the period without actual demarcation of what was testing data.

For the remaining period July 2008 to Dec 2017, the data had intermittent gaps but was judged relatively complete for the indicators on Cattle Prices, Goat Prices, Maize Prices and Mid-Upper Arm Circumference (MUAC). It is for these variables that gap filling was carried out.

- ***Handling missing data***

Even for the socio-economic variables with reasonable observations for most of the years, there are cases of missing observations for some of the years. The options for handling missing data in the study included: ignoring the missing data completely, approximating the missing data or working with the missing data and modelling on probabilities. The study employed the general principles in Higgins, Deeks & Altman (2009) to handle missing data. These principles also have variants in Jeff (2015) and Humphries (2013). The handling of missing data can be summarized into the two main considerations of follow-up with the original data sources for the possible retrieval of any missing data and the decision on the best method of analysis to yield good approximations of the data confirmed to be missing.

Follow up with the original data sources on missing data was made in the cases for which this approach was possible. An example was the case of Turkana county where this follow up lead to extraction of past data from old text-based repositories. The general execution of this process was initiated by the engagement of the data collection units to obtain any data formerly indicated as missing but were in essence in their possession. The change in data completeness over the review period was kept for reference.

The decision on the best analysis method to yield the best estimates of missing data was based on the evaluation of the possible options for handling missing data include the methods like pairwise deletion, single imputation and model-based estimation methods as discussed here next:

- *Pairwise deletion* discards all records with any missing values. Deletions are done even when data is partially missing. The data, therefore, remains with only complete records for the entire period. This was used for some selected broken rows.
- *Single imputation* methods deploy any of the measures of central tendency to estimate the missing values. This approach reduces variability and finally leads to weak correlations. The study avoided this method since it would lead to weak relationships between the data items especially for cases when missing data are outliers like the years with extreme drought effects.
- *Model-based imputation* was the method of choice of the study wherever applicable. The approach replaces missing values with scores predicted from model equations that include regression equations. The study used smoothing splines.

Smoothing splines is a method of fitting a smooth curve to a set of noisy observations using a spline function to approximate the missing data points. A good review of the use of smoothing splines in gap-filling is provided in Musial, Verstraete & Gobron (2011). The underlying assumption is made that the missing data from the monthly data equates noise in the data and that the data if all well collected would fit a smooth curve that is linear after the knots at the endpoints. Smoothing spline, therefore, can be interpreted as some kind of regression line that interpolates the missing values. The gap-filled data, in interpretation, will require an analysis of the impact of the same data. Study findings should keep the impact of the missing data in the analysis, especially in the discussion of the results. A sample application of spline smoothing is shown for MUAC for Turkana county in which we introduced gaps in 6 time periods (months) is shown in Figure 3.14.

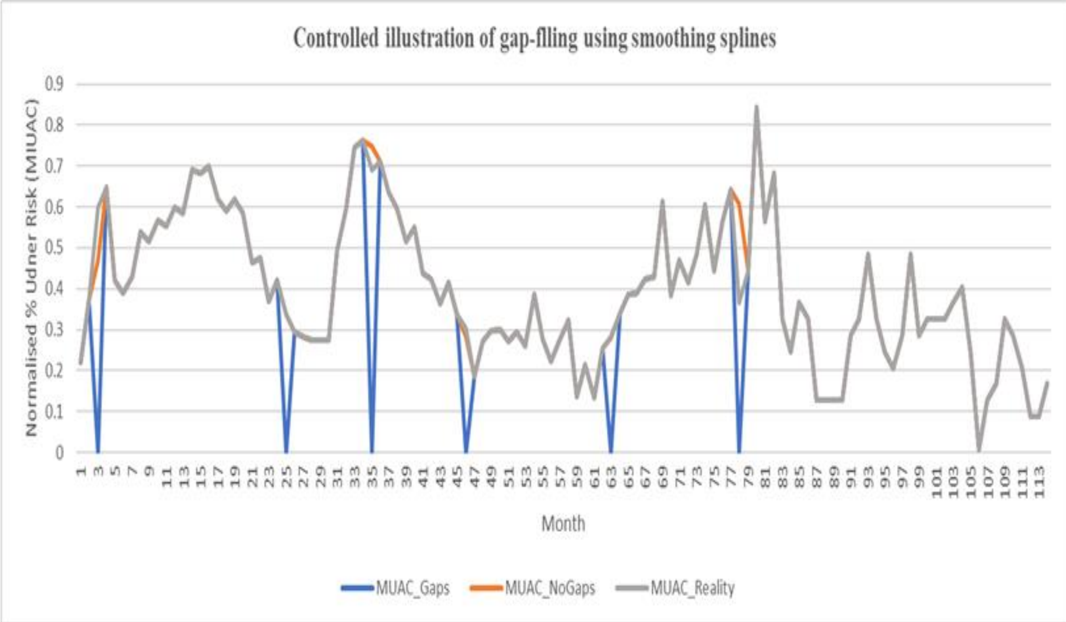


Figure 3.14: Illustration of the use of smoothing splines to gap-fill MUAC.

The case for MUAC in Figure 3.14 illustrates the performance of the Spline smoothing using the three instances of MUAC data. With Gaps (blue line) having the missing data points set to 0(zero), with gaps filled by spline smoothing (orange line) and finally the real MUAC (grey line). It is clear that the use of spline smoothing on the missing data realises a good approximation of missing data and thus good quality gap filling is achieved.

A final step in the handling of the data gaps involved a review of past reports done out of the datasets for specific months with missing data and using these as a verification of the datasets in terms of accuracy. This approach serves to validate the reconstruction done on the data used for the study.

- ***Variable transformations***

  With the data gaps filled, the last step that is key in the data pre-processing steps that we undertook was the transformation of variables. Variable transformation as a tool was used towards two expected benefits. First is the ability of transformations to produce new variables that are better estimators of the variables under study. Secondly, variable transformations also act as a data volumes reduction strategy. One key variable transformation that we did was on the two variables of Goat Price and Maize Price to realize a variable Terms of Trade (ToT). The use of ToT is documented in FEWSNET (2009) to take any of four approaches: (1) food crop to food crop; (2) livestock to cereal; (3) wage to cereal and (4) cash-crop to cereal. We adopt the definition in (2) that calculates ToT out of the relationships between livestock and cereals each case taking the most popular small stock and most popular cereal for the pastoral setting.

The ToT is therefore used to judge the units (Kg) of maize that the local communities get in return for the sale of a goat. The choice of the goat is due to its use by the pastoral communities to meet the cost of immediate and pressing needs of the family. It is expected that as drought impacts the communities, maize becomes expensive as compared to goats and ToT worsens. Lower ToT values thus imply poorer trade conditions for the pastoralists. The ToT is calculated as the ratio of Goat Price to Maize Price. It is therefore unitless and higher ToTs are preferred as it generally indicates better livestock body conditions and hence a proxy for reduced effects of drought on market access for the pastoralists. A key limitation of the ToT is for cases when either of or both the livestock and cereal preference of a unit area is not uniform. This is solved, in our opinion by mapping the different preferences and still calculating the respective ToTs.

3.5.4.2 Presentation of Socio-economic data

The study reviewed the Socio-economic data of the NDMA for production, access and welfare. In general, as presented on the sub-section of completeness in section 3.5.4.1, data collected at the household level, except for MUAC, had many cases of missing data and hence broken time series. MUAC, as a measured quantity was generally robust in the number of records since a minimum number of children were required and even sourced from neighbouring households. The study, therefore, used market access data of both Goats and Maize as represented by Goat prices & Maize prices respectively. MUAC was used to represent utilization. The choice of Goat Price, Maize Price and MUAC were for the following reasons: -

- Completeness of the records for the period July 2008 to Dec 2017 for the three indicators. Goat and Maize prices were generally well collected at the market level from Key Informants. The prices are generally expected to reflect prevailing market conditions as opposed to household levels where panic sales and sales as a result of isolated economic pressures that could be unrelated to drought effects do happen.
- Consistency in the collection of the data especially for MUAC that is a measured quantity. This has ensured the availability of objective records on malnutrition.

The socio-economic data for this study are therefore presented as a set of two variables: Terms of Trade (ToT) and MUAC. The modelling uses lagged values of these variables to predict future nutrition conditions. The variables are used in combination with those identified to be predictive of drought severity.

3.5.5 Summary of the study variables

This is an observational research study using archival data from existing data archives. The approach involves the analysis of co-variant data to determine pre-existing relationships without any attempt to manipulate the predictor variables. The chosen research method is therefore driven by statistical analysis in the investigation of relationships between different datasets. The study, therefore, infers correlations rather than causality. The definition of correlational research is based on the data analysis methods employed on the data rather than the data gathering method. The typical outputs of correlational analysis are Scatter Plots, Regression line, Correlation Coefficient (r) and Coefficient of determination (R^2). The variables in this study are therefore grouped into either predictor variables or predicted variables.

The dataset integrated from all the sources is highly multivariate. The case of multiple datasets quantifying the same concept necessitated variable selection among such, especially for competing datasets. The predicted variables were carefully defined to correlate with drought severity and subsequently with drought effects on nutrition. For this study, we refer to the variables used to predict the drought severity and drought effects as the *predictor variables* rather than as the *dependent variable* that implies causality and thus fits only within experimental studies. Since we use correlational design for the study, we use the term *predicted* or *target* variable to refer to the variable that defines the drought status.

The definition of the predictor variables is influenced by the fact that the use of the correlational research design implies that there is a measuring of variables and assessing the relationships between them without manipulating the independent variables. This is as opposed to the case in experimental research where the measurements indicate the effectiveness of treatments done on manipulated aspects of a study. For each level of investigation, causality is not inferred in correlational studies as would be the case in experimental research.

The variables used in the study are either remote sensing variables or socio-economic variables. The variables, as discussed in chapter 3 are summarized in Table 10.

Table 10: Summary of the study variables

Type of drought	Variable & Transformations	Description
Meteorological	Rainfall Estimates (RFE1M & RFE 3M)	The absolute value of Precipitation (mm) as recorded from Satellites
	Standardised Precipitation Index (SPI1M & SPI3M)	Standardized values of Precipitation aggregated over the 1M & 3M time periods respectively.
	Rainfall Condition Index- RCI1M & RCI3M	Normalized values of Precipitation aggregated over the 1M & 3M time periods respectively
Hydrological	Land Surface Temperature LST1M & TCI1M	Land Surface Temperature & Emissivity data Aggreged at 1M for both temperature and temperature condition index (TCI) calculate as RCI above.
	Evapotranspiration (VET) EVT1M	EVT, the sum of EV & T is, used to calculate regional water and energy balance, soil water status and applied especially in water resource management.
	Potential Evapotranspiration (PET) PET1M	PET is the demand side of water from the atmosphere. Aggregated over 1M
	Standardized Precipitation & Evapotranspiration Index (SPEI) SPE1M & SPEI3M	Based on climatic data. It is a standardization of the equation $Di = P_i - PET_i$ aggregated over 1 & 3 months respectively
Agricultural	Normalized Difference Vegetation Index (NDVI) NDVIDekad	NDVI value for the last ten days of the period of prediction. Rage is -1 to +1 indicating the photosynthetic vibrancy of vegetation.
	Vegetation Condition Index (VCI1M & VCI3M)	Normalized values of NDVI over the 1M & 3M time periods respectively. Aggregated over 1M and 3M periods.
Socio-Economic	Goat Price	Access variable- the monthly average of goat prices
	Maize Price	Access variable- the monthly average of maize price
	Terms of Trade	Kilograms of maize exchanging for a goat
	MUAC	Percentage of Children under 5 years with MUAC<135

3.6 Modelling Approach

This section corresponds to the section on data analysis and data modelling as outlined in Figure 3.4 on the logical flow of the study but with an initial description of how the problem that overarches data analysis and modelling is defined.

The modelling approach for this study is informed by the elements of the review of literature that was carried out based on the problem statement. The approach is thus based on appropriateness for the purpose that is informed by the following key areas:

- The correlational research process elaborated in Figure 3.3 has the Analysis phase as the last phase of correlational design. This we expanded to include both presentation and interpretation of the results of the analysis.
- The CRISP-DM Methodology for data mining in Figure 2.11 has in its approach the steps for modelling and evaluation with the techniques for the stage chosen based on appropriateness for data. This approach advocates the use of a separate dataset for model evaluation and views evaluation as a distinct process to modelling.
- The SEMMA methodology reviewed in section 2.8.2.4 provided for the steps to Model and to Assess. These are similar to the *modelling* and *evaluation* steps in CRISP-DM and the analysis phase in the design of investigation presented earlier in section 3.4.1.

The methodology of this study followed on the above to define the steps for the study post data collection as based on the steps outlined in the process adopted for the study that is defined in the Figure 3.15.

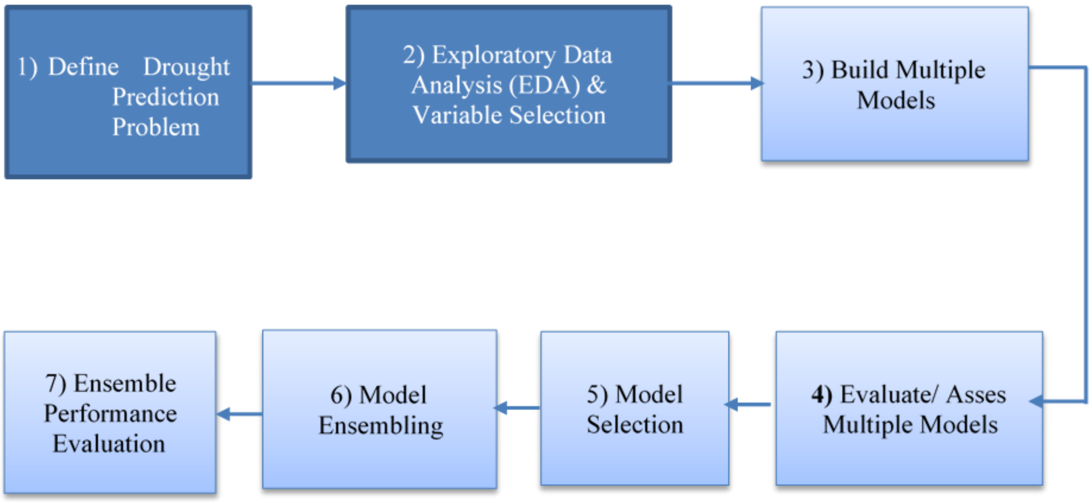


Figure 3.15: The model building process post data acquisition.
The model building process is based on the building of multiple models, their validation, ensembling and subsequent performance evaluation.

We, here next, discuss the approach employed by the study following the above steps of the modelling methodology.

3.6.1 Methodical definition of the problem

To be able to develop predictive drought monitoring models, the phenomenon of drought needs to be well defined. From the literature review and with the key concepts in the definition of drought, we formulate the drought prediction problem using Equation 20.

$$D_{(i,j)} = f(x_1, x_2, x_3 \cdots x_n) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (20)$$

Where $D_{(i,j)}$ is a quantification of drought severity for a spatial extent *i* at time *j*, f is a function that accepts a set of n (n ≥1) variables and transforms them to approximate the real $D_{(i,j)}$. The *n* variables $x_1, x_2, x_3 \cdots x_n$ are predictor variables that are used to monitor drought from meteorological, hydrological to agricultural drought.

The relationship between the estimator variables and drought severity in Equation 20 is not necessarily linear just as the *n* variables are not guaranteed to be normally distributed. The function f is assumed to have a process for selecting the best subset from the *n* variables that best approximate the drought severity.

The socio-economic variables are essentially effects of drought and are therefore measures of the effect of $D_{(i,j)}$on the social and economic aspects to the exposed elements at spatial extent ***i*** at time ***j.*** This implies that the definition of the problem should be expanded to include the modelling to approximate the effects E of drought. This takes the same definition as that of drought severity above and is formulated in Equation 21.

$$E_{(i,j)} = g(x_1, x_2, x_3 \cdots x_n) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (21)$$

Where $E_{(i,j)}$ is a quantification of drought effects for a spatial extent *i* at time j, g is a function that accepts a set of n (n ≥1) variables and transforms them to approximate the real $E_{(i,j)}$. The *n* variables $x_1, x_2, x_3 \cdots x_n$ are predictor variables that are used to monitor drought severity from meteorological, hydrological, agricultural combined with the variables for socio-economic drought.

The variables $x_1, x_2, x_3 \cdots x_n$ are therefore interpreted to be the set of variables used for the monitoring of meteorological, hydrological and agricultural drought severity for $D_{(i,j)}$. On the other hand, for $E_{(i,j)}$ we use the combination of variables used in the approximation of $D_{(i,j)}$and those variables identified for measuring the socio-economic effects of drought. The modelling task is therefore meant to approximate both $D_{(i,j)}$and $E_{(i,j)}$ based on the choice of suitable techniques informed on by both the variables and the data identified for the study.

In the study datasets, the definition of the two target variables $D_{(i,j)}$ and $E_{(i,j)}$ was done based on: VCI aggregated over 3 months period (VCI3M) and the proportion of children at risk of malnutrition as defined by the variable MUAC. The use of the VCI to define drought severity- $D_{(i,j)}$ was based on four key reasons as documented in Adede et al. (2019b):

- with a range of 0 to 100, VCI is easy to interpret.
- as an index, VCI is indicative of agricultural drought, which is a later stage drought as compared to meteorological drought indicated by SPI.
- VCI is more directly related to food and fodder availability in the study area compared to SPI.

- VCI is a measured quantity as opposed to the SPI that is, for the case of this study, a modelled quantity

The second target variable on drought effects- $E_{(i,j)}$ was defined based on MUAC conditions. Similar to the definition of drought severity, MUAC was chosen to define drought effects for three key reasons:

- MUAC is a measure of late-stage socio-economic drought as compared to the other indicators derived at the market level (Figure 2.9). This is so because MUAC is indicative of drought effects on malnutrition and in the case of the datasets used is defined as malnutrition of children aged below 59 months.
- MUAC is physically measured at community level from children from multiple households but recorded in the household questionnaire. This is as opposed to the prices that are based on Key Informants and are thus based on recall. We deem MUAC more of a reliable indicator than prices in this respect.

With the prediction problem well defined for both drought severity and drought effects, the study then proceeded in two parts. First is the investigation of how the different combinations of the variables $(x_1, x_2, x_3 \cdots x_n)$ estimate $D_{(i,j)}$. The second is to investigate how the best estimators of $D_{(i,j)}$ combined with socio-economic variables approximate $E_{(i,j)}$. A further discussion on the methods to accomplish these tasks is presented in section 3.6.3.

Prior to the discussion on the model methodology, the presentation of the data for prediction also becomes a big question in the context of this study. Given we have observational data, we present the data for drought monitoring as a lag of the variables that define drought severity based on the lag of its predictors. To predict drought severity, we lag the variables and then use them to predict future values. In essence, all the predictor variables $x_1, x_2, x_3 \cdots x_n$ are presented as lagged by one month for each $D_{(i,j)}$and $E_{(i,j)}$ in the training dataset. The lagging of the variables, therefore, gives an opportunity for the use past lags of the target variables in the prediction of future values.

3.6.2 Exploratory data analysis

In the exploratory data analysis process, we investigate the relationships both amongst the predictor variables $(x_1, x_2, x_3 \cdots x_n)$ and between the predictors and the target variables of both drought severity $D_{(i,j)}$ and $E_{(i,j)}$. This analysis culminates with the selection of the subset of the variables $(x_1, x_2, x_3 \cdots x_n)$ from the set of all variables considered predictive of the target variables.

Since the study is based on correlational research design using observational data, we do the exploratory data analysis using statistical analysis methods which are particularly appropriate for cases where a dataset has multiple variables. The variables in this study are all ratio data and therefore support a raft of methods of analysis. The study, therefore, uses both descriptive and inferential statistics. We use descriptive statistics for numerical summarization of data and inferential statistics to infer relationships between variables. The suitability of statistical analysis is for studies that aim to summarize descriptive data or that aim to understand relationships between different variables to offer possibilities of generalizations. The study chose to first investigate the methods that were appropriate for the datasets that were pre-processed. The appropriate methods were then chosen from the set of both parametric and non-parametric methods. The non-parametric methods investigated and chosen from included Spearman Rank Correlation (SRC), Wilcoxon Test, Man-Whitney, Kruskal-Wallis and Friedman tests.

The ultimate aim of the exploratory data analysis phase is the selection of variables to be used in the modelling process. In this study, we aim to build multiple models. Variable selection is therefore reduced to the decision of decreasing the number of data sources, especially for similar data. Exploratory data analysis was thus applied in choosing which between TAMSAT and CHIRPS rainfall data was to be used in the modelling process. The selection between TAMSAT and CHIRPS was done using several methods and a final inference drawn based on the multiplicity of the methods.

For variable selection, normality or otherwise was proven for the indicators, information criterion established using Akaike information criterion (AIC), the relative importance of variables was undertaken as well as stepwise regression and a modelling approach to variable selection. Agreement and inference between these multiple methods were used to settle on the set between TAMSAT and CHIRPS to maintain in the modelling process.

Spearman Rank Correlation (SRC) analysis was also done. SRC tests the direction and strength of the relationship between the two datasets under investigation for relationships. SRC, therefore, tests if there is a relationship in the way two sets of observations vary. While the *Wilcoxon Test* compares two paired sets of data, calculates the differences between each set of pairs and analyzes the list of differences; the *Man-Whitney* tests whether two sets of observations come from the same distribution, with the Null hypothesis set as the probability from one population being higher than that of another. The *Kruskal- Wallis test*, however, tests the equality of population medians among groups using one-way ANOVA by ranks and differs from the *Friedman test* that compares three or more paired sets of data**.**

3.6.3 Modelling methodology

With the problem defined in section 3.6.1, the modelling methodology is reduced to the search for all the *f's* that approximate both the drought severity ($D_{(i,j)}$) and all the *g's* that approximate drought effects ($E_{(i,j)}$). The definition of the problem mirrors that of a machine learning and is adopted as the search for all the functions (*f's & g's*) from the space of all possible functions (F_U) that approximate ($D_{(i,j)}$) and drought effects ($E_{(i,j)}$) with some degree of accuracy as measured by some pre-selected measures of performance (P).

From Figure 3.15 on the study modelling methodology, model building methodology covers the steps 3-7 discussed here next.

- ***Model building***

To precisely formulate the methodology for the building of models, the following key concepts of variable space, model space (F_U), model and modelling technique need a formal definition.

- ***The variable space*** is the set of all variables $(x_1, x_2, x_3 \cdots x_n)$ that are deemed capable of measuring both ($D_{(i,j)}$) and ($E_{(i,j)}$). It corresponds to all the variables identified in the conceptual framework for this study.
- ***The model space (F_U)*** is the set of all possible models that can be derived from all the possible combination of the variables deemed to measure both ($D_{(i,j)}$) and ($E_{(i,j)}$). The model space is interpreted to imply the set of models that can be built using a machine learning technique (M) with each model defined as a combination of the variables deemed to measure drought severity or drought effects.
- ***The machine learning technique*** $(\boldsymbol{M})$ is defined based on literature to be any of the methods for approximating prediction functions f based on historical data used in the training of models. The formulation of the drought forecasting problem as indicated above reduces it to a machine learning problem. In the study, we define *f* as a supervised machine learning regression function and aim for output in the [0,1] range with the ability to be stretched to the [0,100] range for ease of interpretation. In this study (M) includes the study case techniques of Artificial Neural Networks (ANN) and Support Vector Regression (SVR) for model building. The statistical approach of General Additive Models (GAM) are also used but more as a basis for variable selection rather than the approximation of all the *f's* that are ultimately ensembled. The model building techniques of choice are elaborately discussed in section 3.6.4.
- ***The model f*** is, therefore, a single function that outlines the combination of a subset of the variables from the variable space used to approximate drought severity and effects. In other words, $f \in F_u$.

With the modelling methods chosen as ANN and SVR, the task is to get the set of all *f* that can be learnt using both techniques. The process for building the models given that they will be ranked against each other and a selection done will follow on the use of the same dataset for both the training and evaluation of the models. The study follows the 70:30 split of data during the model training phase and subsequently keeps aside data for model testing. This approach is referred to as the 3-way data split and is as illustrated in Figure 3.16. For a better approximation of model performance, this random split is done 10 times on the training dataset and bootstrap samples passed to the models whose performance in both training and validation across the sets is then averaged.

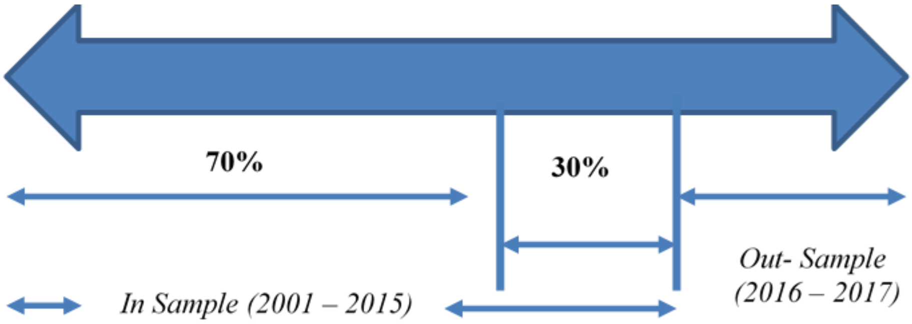


Figure 3.16: The in-sample (70:30) approach to model building and validation with a separate dataset for model testing.

Given the number of variables and their possible combinations to form models, the model space (F_U) is expected to be quite a big model space. A visualization of the number of models given the number of variables per model is as shown in Figure 3.17 for an illustrative case of 20 variables.

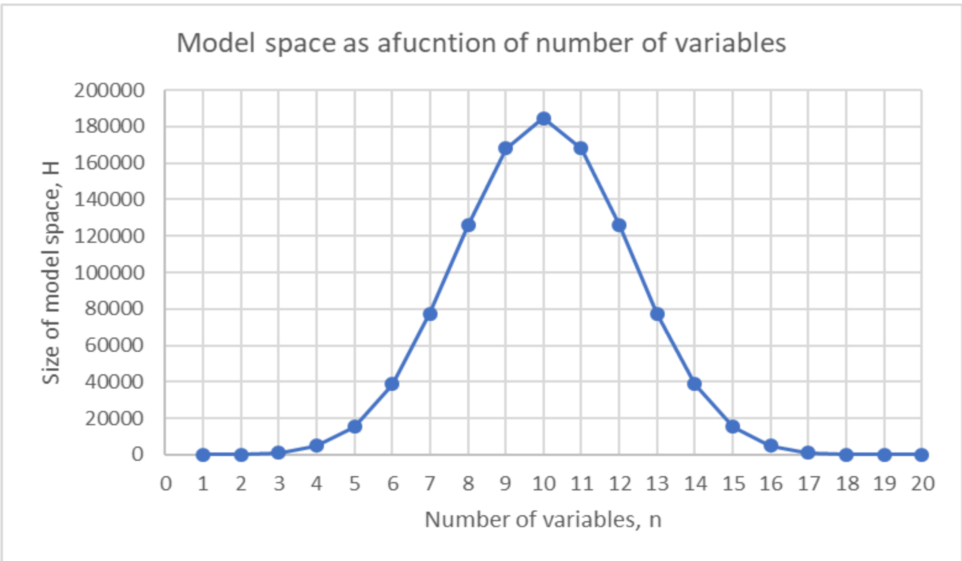


Figure 3.17: Model space a function as a function of the number of variables.

Given a modest set of 20 predictors, a total of around 1.05 million models would need to be built. The complexity of the model spaces given the number of variables in this study would make building all the models an exercise in futility. Building all the models, for their sheer number, would be akin to a learning task without a priori assumptions that would be contrary to the no free lunch theorem (Wolpert, 2012) as applied to supervised learning. A set of assumptions are therefore enforced to reduce the model space. These assumptions are as defined below:

- There is no value in the use of more than one variable defining the same phase of drought in the same model.
- That variables are better separated into those on drought severity and those on drought effects with those on drought severity also separated into the drought types.
- That different lags of the same variable increase complexity in prediction at no extra gain in prediction and ease of explanation of the models.

The effects of making these assumptions are later discussed as part of the results of this study and are shown to massively reduce the modelling space making it possible to reason on.

- ***Model evaluation/ assessment***

Each model (*f*) realised using each of the modelling techniques ($m \in M = \{ANN, SVM\}$) needed to be evaluated for their performance in prediction. Performance assessment was done using a standard set of criteria including determinant of correlation (R^2), mean absolute error (MAE), mean standard error (MSE), root mean squared error (RMSE), mean absolute percentage error (MAPE). For objectivity in the evaluation of model performance, the three data split approach is used. Models are built and validated on the 70:30 split of the training datasets. Due to the effect of factors like climate change, there is a tendency of the future not to be exactly characterised as the past. We ensure this possible challenge to internal validity is handled by keeping aside two years of data for model testing. The two years of data kept aside for model testing is referred to as the out of sample dataset or testing dataset.

The chosen method of model performance evaluation was the determinant of correlation- R^2. In the case of ties, the other measures were considered as tie-breakers. It is, however, to be noted that we considered the limitations of R^2 in our evaluation methodology and opted for Adjusted R^2 as the basis of the evaluation of model performance. As indicated in Equation 22 and Equation 23, R^2 is basically a function of RMSE and is by extension closely related to the other measure of performance.

$$RMSE = \sqrt{MSE} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (22)$$

$$R^2 = 1 - \frac{SSE}{TSS} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ (23)$$

The numerator of Equation 3 is the sum of squared error (SSE) which is, in essence, the expression $n \times MSE$ if sum errors for n estimations. R^2 can thus be expressed in terms of MSE as in Equation 24

$$R^2 = 1 - \frac{n*MSE}{TSS} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ (24)$$

The implication of Equation 20 is that holding the TSS constant implies an inverse relationship between R2 and RMSE. It is for this reason we use adjusted R2 in combination with other measures of error.

Further justification for the use of Adjusted R^2 apart from its proportionality to the measures of error (MSE, RMSE etc) are as follows: -

- Just like R^2, Adjusted R^2 has the property that it scales nicely between 0 and 1 (can be interpreted as between 0 & 100%)
- R^2 is noted to rise with the increase in the number of predictors even when the increase is artificial and does not improve model's fit, Adjusted R^2 is noted to incorporate the model's degrees of freedom. Adjusted R^2 is, therefore, most appropriate for multiple predictor models.

- Measures the proportion of total variance that is explained by the model
- Best for evaluating prediction between variables as opposed to the relationship between variables.

- ***Model selection***

The methodology of this study is much biased towards model selection more than variable selection. The intent is to reduce the space of models that are used in the model ensembling process. Due to the possible number of models in the model space, careful crafting had to be undertaken to select models objectively. The selection of models has to happen in an automated way using the pre-selected performance metrics. We do model selection at different stages. First, using the GAM modelling process to reduce the model space. Thereafter, we build models of both ANN and SVR that we subsequently select for model ensembling. All stages of model selection are based on models with $R^2 \geq 0.7$ with ties broken using any of RMSE, MSE and MAE in that order.

- ***Model Ensembling***

We investigated two approaches to model ensembling using three methods of model ensembling. First, we ensemble the models using three distinct methods. These include linear averaging, rank weighted averaging and ANN randomly learnt weight averaging. The model weighting methods are used to realised either homogeneous or heterogeneous model ensembles. We define homogeneous model ensembles as model ensembles of models learnt using the same modelling technique like ANN or SVR in the case of this study. On the other hand, heterogeneous model ensembles are the ensembles where the bases models are built using different techniques. In this study, the combination of both the ANN and SVR base models in the model ensembles, therefore, defines heterogeneous model ensembling.

The model ensembling method is generally defined based on the three-step process outlined as: -

- Generate the experts/models that will be used
- Score using each expert separately
- Combine the experts and "average" their values

In this study, the ensembling process is implemented using the three approaches as presented in Figure 3.18 and described thereafter in the order of their complexity in implementation.

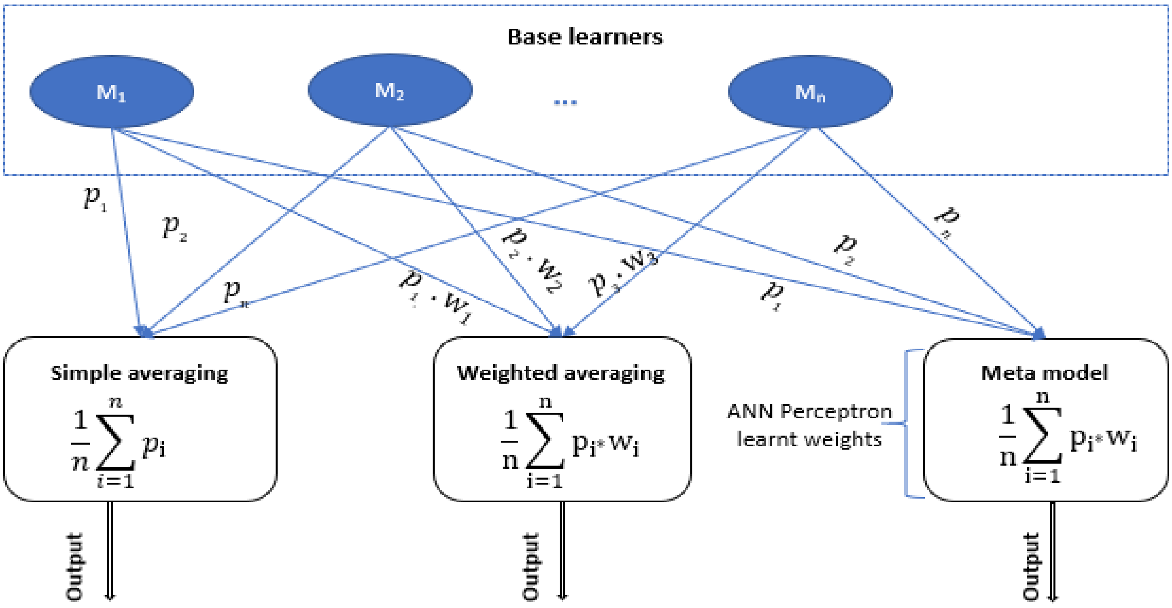


Figure 3.18: Schema of the model ensemble approaches.

The approaches in Figure 3.18 are simple averaging (left), weighted averaging (centre) and model stacking (right).

The implementation of the ensembling approaches is as follows:

- ***The simple average model*** is based on running all the models on the test data and making predictions from every one of them. The outputs of all the models are averaged according to Equation 25. This method is by and large a democracy of equal voting rights without the ability to discriminate models based on their performance as based on the performance metric already used in the evaluation of the model performances.

$$\frac{1}{n}\sum_{i=1}^{n} p_i \text{ ………………..…. } (25)$$

With p_i as the prediction from the ith model.

- ***The weighted average model*** just like the simple averaging ensemble method is based on running all the models on the test data and making predictions from every one of them. The outputs of all the models are averaged according to Equation 26. This method biases the weights in the ensembling with good performer models given higher weights in the ensembles. In this study, the weights were based on normalized values of model performance as assessed by R^2.

$$\frac{1}{n}\sum_{i=1}^{n} p_{i^*} w_i \ldots\ldots\ldots\ldots\ldots \quad (26)$$

  Where w_i is the normalized value for each model in the ensemble such that all the weights sum to 1(one). The weights are therefore stretched between 0 and 1 centred around the minimum with the maximum value as the scale.

- ***The stacked model approach*** is a more complex implementation that assumes that the future performance of the models is not known before-hand. In this approach, a random set of weights is leant for each modelling technique individually on the validation dataset. The outputs of each model become the features for which weights are learnt using a new modelling technique. In this study, a perceptron neural network is randomly built to optimize the estimation of the target variable using the outputs of the base models. The set of weights that best approximates the output are chosen as the random model weights that are then used for the ensembles. This method has a single limitation of raising the possibility of the curse of dimensionality as it scales very fast in features with the increase in the number of ensemble members.

The estimation of model performance, especially in the case of model ensembles is particularly not a naïve process. The performance of the ensemble models on the test dataset is directly measured using the same performance metrics as that of the member models. R^2 remained the basis of analysing the performance of the ensembles as was the case for the base models. The resulting ensembles are then compared in performance to the best performer models that are referred to as Champion models.

3.6.4 Model building techniques

The methodology of this study is to use a statistical method as an initial modelling method that is used, essentially for the selection of variables, and the subsequent development of the predictive models using a machine learning method. We use General Additive Models (GAM) for variable selection and Artificial Neural Networks (ANN) for the development of the predictive models in the pre-study. The main study used the Support Vector Regression (SVR) technique in the model building process in addition to the ANN technique. These model building techniques are described here-under.

- ***Generalized Additive Models (GAM)***

GAM models were selected because they do not assume linearity between the predictor and the response variables (Hastie, 2017). In addition, GAMs are free-form since they do not require the ascertainment of the functional form of relationship to be modelled beforehand. In the case that the relationships are best approximated by linear, quadratic or cubic functions, GAM results simplify to these as is appropriate. These coupled with the fact that we still have the desirable features of GLiMs and GLMM make GAM models a viable tool for weather-based data modelling. GAM models are expressed as shown in Equation 27.

$$Y = a + f_1(x_1) + f_2(x_2) + \cdots + f_n(x_n) + \varepsilon \ldots\ldots (27)$$

where a is an intercept and f are smooth functions; Y is the response function and x_1 to x_n are the n predictor variables. Smoothing functions are either local linear regression or splines. In practical application, caution is advised since smoothing in GAM generally leads to model overfitting.

The space of models for the study, as given by Equation 28, is around 2.15 billion. This space would be impractical to navigate in the search for the best predictor model.

$$\sum_{r=1}^{31} \frac{n!}{(n-r)!\,r!} \approx 2.15 \text{ billion} \ldots\ldots\ldots\ldots\ldots (28)$$

To minimise the space complexity, we make some *a priori* assumptions to avoid the futility of bias-free learning and also follows Occam's razor (Mitchell, 1997). First, we assume that a maximum of two variables in the GAM models will give us reasonably simple models that remain predictive. Second, we assume that including multiple variables of the same drought type is both an unnecessary increase in the complexity of the model space at a marginal possible increase in performance. Together with these two assumptions, we further use a rule of thumb to not use different lag times of the same variable in a single model. To capture seasonality, we further included an additional variable for the month of the year of the instances as a sine wave in the GAMs. Seasonality is expected to exist in vegetation cover data apart from the standardised and relative range variables. The most seasonality prone datasets are the vegetation variables that use absolute values.

With the model space reduced to 102 from the possible initial space of over a billion models, we brute-forced the process of training and evaluating the models in an automated process. Multiple model evaluation metrics were used and the results logged for both model training and model evaluation.

- ***Artificial Neural Networks***

The choice of ANN for the study was most informed by their ability to do regression and much more the ability to handle cases of non-linearly separable training data. They, however, have the disbenefit of complexity in interpretation and are thus sort of a black-box approach to learning.

To overcome overfitting which is the most common limitation of ANNs, we chose models that are judged to perform better in the evaluation datasets as compared to the training datasets using R^2 as the measure of model performance. We formulated a working definition of model overfitting. The ANNs were built on normalized variables. Variable normalization was done before model training was therefore done for both the training and validation datasets. The variables were all normalized to the [0,1] range. The values were centred at the minimum value for each variable then linearly scaled between the minimum and maximum values.

The ANNs were built using a variant of the back-propagation algorithm- the resilient backpropagation (RPROP) as proposed in Riedmiller (1994). RPROP is considered to side-step the hyper-parameter tuning problem. For the question on the complexity of configuration, the modelling process was set to have a formation of 2-5-3-1 and thus had two hidden layers. The configuration was realized following on a rule of thumb process sequenced as follows: -

- Since the data is expected to exhibit non-linearity, two layers would achieve to learn any arbitrary function. (a-b-c-d).
- Settle on the input models based on the number of input variables. This was set at a maximum of 3 (3-b-c-d)
- Fix the output parameters for drought severity or drought effects. This is 1for each case (3-b-c-1)
- The input layers must have at least 2 nodes so as not to be a linear transfer of weights. So, the minimum has to be 3-2-2-1.
- Experiment adding one node to the hidden layers left to right till the ANN converges or any other special condition is met. In this case, we set the special condition to be that at least 50% of the models were deemed predictive.

The formation was realised to ensure convergence of all models within the model space. The actual run of the process, therefore, had the configuration of 5,3 for the two hidden layers making a total of 8 hidden neurons. The maximum step was set to 1e+06 and this represented the maximum steps for the training of the neural network at whose attainment the neural network's training process is stopped. The maximum step size was a failsafe condition for the ANNs, should the pre-selected set of hidden layers not lead to convergence majorly due to partitions in the training and validation datasets.

- ***Support Vector Regression***

Support Vector Machines (SVR), just like the Support Vector Machine (SVM) is a Kernel-based machine learning approach that is non-parameterized and thus does not demand assumptions on the distribution of the training data. For this study, we used the support vector regression approach that though related to SVM, defines the case for prediction of real-valued outputs. A comprehensive review of the SVM approach is documented together with evidence of its superior performance in classification in Wagacha (2003) and Mountrakis, Im & Ogole (2011).

In our approach, we take a formulation of the SVR in which: -

- The training data is linearly inseparable. This implies we use non-linear kernels
- The data has noise and so perfect fit is not ultimately achievable. This means we will opt for a soft margin.
- We retain control of how much error *(ϵ)* we tolerate for the model. However, we penalised any error above*(ϵ)* by a cost parameter *C.*
- We do not control the number of support vectors

We, therefore, define the problem of SVR learning based on the model defined in Equation 29. The kernel is defined as a sigmoid kernel akin to the definition adopted in artificial neural networks. The definition of gamma as the inverse of the dimension of the data is adopted.

$$Model_{svm} = svm(f, d, k, \epsilon, C, g) \dots \dots . \dots (29)$$

Where *f* is the model formula defining the target variable in-terms of the features, d is the corresponding training dataset, C the cost of incorrect prediction as defined by the insensitive-loss function ϵ.

In summary, we present the modelling methodology as shown in Figure 3.19 outlining the key decision points in the modelling methodology. The use of different facets of the modelling methods including the bugging of datasets and ensembling of models are deployed to realize improved model performance.

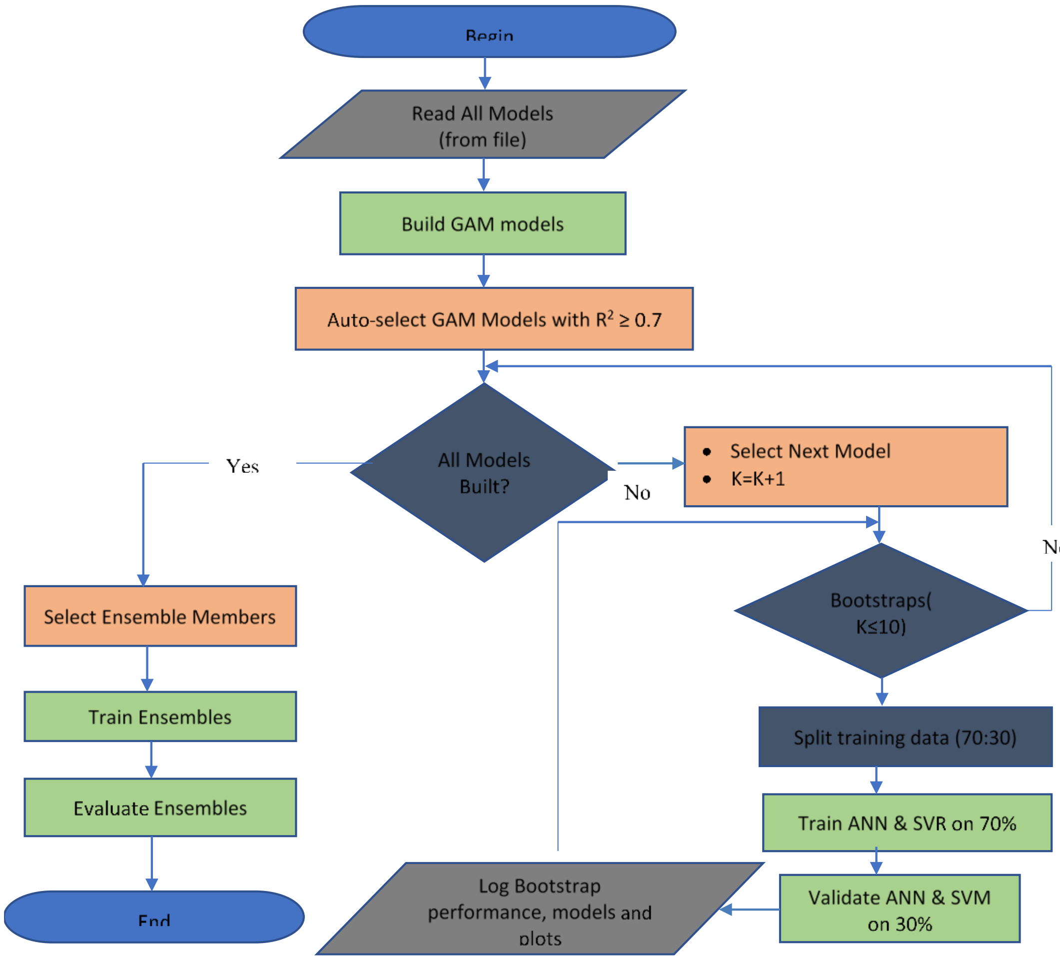


Figure 3.19: Outline of the modelling process.

The modelling process sequentially inputs the models realized after assumptions through the GAM modelling process. Models with $R^2 \geq 0.7$ were then subjected to a bagged modelling process using both ANN and SVR on similar datasets for both training and evaluation. The data was randomly partitioned in the ratio 70:30 for training and validation respectively. The 70:30 split was for each and every iteration of the k times data was used to build and assess the models. The same training data was passed across all the models to unsure outputs are comparable.

3.7 Pre-study Methods & Results

To ensure success in the study and the achievement of the study objectives, we did a pre-study. The pre-study involved the use of both precipitation data and vegetation indices to realise multiple drought prediction models and to subsequently use a set of model selection criteria to reduce the complexity of the model space.

The objectives of the pre-study are summarized as follows: -

- Establish the appropriateness of methods, especially of the proposed general additive model (GAM) process for model selection.
- Investigate and answer the question on the most appropriate lag-times on the data to use that would also form the basis for the period ahead to be predicted using the identified variables. The basis is to realise the maximum lag time in the future which the prediction needs to consider.
- Validate the assumption of non-significance of the gain in performance of models realized from the use of multiple lags of the same predictor variables
- Validate the assumption that pairing of vegetation and precipitation datasets avoids the problem of multi-collinearity.
- Document the performance limitation of the traditional approach of identifying a single champion model particularly in terms of performance and also in the prediction of outliers.

In this section, we provide the results of the pre-study as documented towards and from the contents of the first publication out of the research work as documented in Adede et al (2019a).

3.7.1 The problem of drought in the study area

The study analyzed past droughts to form the basis of highlighting the problem of droughts in the study area. This offers the evidence necessary to judge the usefulness of the study.

Klisch & Atzberger (2016) documents the Vegetation Condition Index (VCI) as a temporal and spatially aggregated anomaly of the Normalized Difference Vegetation Index (NDVI). While the NDVI gives absolute vegetation status for a given spatial extent at any given time, the VCI scales the actual NDVI value in the range between a historical minimum (VCI = 0%) and maximum (VCI = 100%) for the given time unit.

The classification of drought based on the 3-monthly aggregated Vegetation Condition Index (VCI) following on the thresholds used variously in Klisch & Atzberger (2016), Meroni et al. (2019), Klisch, Atzberger & Luminari (2015) and in Adede et al. (2019a) over the 178 months covering March 2001 to December 2015 shows in Table 11 why the problem of droughts is quite immense for the study area.

Table 11: Summary of monthly drought phases for the counties in the study region (03-2001 to 12-2015)

County	Extreme	Severe	Moderate	Combined
Mandera	8	31	43	82
Marsabit	8	26	70	104
Turkana	4	28	64	96
Wajir	9	25	61	95
Total	29	110	238	377

Over the 178 months (Table 11), 377 out of a possible 712 (52%) drought episodes were reported at the county level, 29 (4%) of which being classified as extreme (VCI<10) and therefore signalling the possible collapse of community coping capabilities. A drought early warning system with predictive capabilities is thus a possible value addition to the drought monitoring process for the counties in this study area.

This problem of drought particularly gets exacerbated when the poverty headcount for the counties in the study area is factored in. The poverty rates across these counties are above 60% with Turkana (79.4), Mandera (77.6), Marsabit (63.7) and Wajir (62.6) as documented in KNBS (2018). These high poverty rates make the effects of droughts even more impactful on the communities due to the already existing high vulnerabilities.

3.7.2 Pre-study Data

The variables used in this predictive study comprised both precipitation and vegetation indices with either 1 or 3-month aggregation periods. The precipitation datasets were derived from TAMSAT (Tarnavsky et al., 2014) and include Rainfall Estimates (RFE), Rainfall Condition Index (RCI) and the Standardized Precipitation Index (SPI). Vegetation conditions were characterized through the Normalized Difference Vegetation Index (NDVI) and Vegetation Condition Index (VCI) directly provided by BOKU (Klisch & Atzberger, 2016). The above precipitation and vegetation indices were calculated at pixel level then aggregated over the appropriate time-scales and administrative boundaries. The details and formulae for the computation of these indices are as provided in Table 12.

Table 12: A description of the index calculation formulas

Variable /Index	Index Calculation	Index Description
NDVI	NDVI = (NIR – Red) / (NIR +Red)	Predictor variable; measures the average monthly vegetation greenness
VCI	VCIc,i=100* (NDVIc,i- NDVImin c,i)/ (NDVImax c,i- NDVImin c,i) (Klisch & Atzberger, 2016)	Predicted variable aggerated over 3 months period.
RFE	Rainfall estimate from TAMSAT product (in mm) (Tarnavsky et al., 2014)	Predictor variable; an estimate of the monthly rainfall
RCI	RCIc,i=100* (RFEc,i- RFEmin c,i)/ (RFEmax c,i- RFEmin c,i) (Du et al., 2013)	Predictor variable: [0,1] normalized RFE for each extent and for each time period
SPI	For each location, c and period i, the long-term record of TAMSAT RFE was fitted to a probability distribution then transformed to a normal distribution so that SPImean c,i=0 (WMO, 2012)	Predictor variable; standardised RFE for each extent and for each time period

In Table 12, $NDVI_i$ indicates the NDVI observed at time i; NDVImin and NDVImax are minimum and maximum NDVI observed in the period 2003-2013. NIR and Red are the spectral reflectances in near-infrared and red spectral channels of MODIS satellite, respectively. Before use, the NDVI time series is smoothed and filtered to remove the negative impacts of poor atmospheric conditions and undetected clouds (Klisch & Atzberger, 2016).

Klisch & Atzberger (2016) and Meroni et al. (2019) document the use of Vegetation Condition Index (VCI) as a temporal and spatially aggregated anomaly of the Normalized Difference Vegetation Index (NDVI). While the NDVI gives absolute vegetation status for a given spatial extent at a given time, the VCI scales the actual NDVI value in the range between a historical minimum (VCI = 0%) and maximum (VCI = 100%) for a given time unit. The widely used time units in the calculation of the indices are dekads which are 10-day periods, months and seasons which is mostly 3 months. The choice to use the BOKU dataset for VCI is based on the fact that the same is deployed in Kenya's operational drought monitoring system at the NDMA. The data is sourced from the NDMA monitoring systems as deployed by BOKU. Several studies have evaluated the BOKU dataset including comparisons against similar products (Atzberger et al., 2016; Jensen et al., 2019; Klisch, Atzberger & Luminari, 2015 and Meroni et al., 2019).

The indices in Table 12 were further subjected to a variable transformation step that resulted in the variables used for drought modelling. All the variables as provided in Table 13 are lagged predictors with the predicted variable that is provided in ***bold*** (VCI3M) as the only non-lagged variable. The month of the year was added to model seasonality.

Table 13: Variables used for modelling

Index	Variable description	1-Month Lag	2-Month Lag	3-Month Lag
NDVIDekad	NDVI for the last dekad of month	☒	☒	☒
VCIDekad	VCI for the last dekad of month	☒	☒	☒
VCI1M	VCI aggregated over 1 month	☒	☒	☒
VCI3M	VCI aggregated over the last 3 months. The non-lagged value is the dependent variable	☒	☒	☒
RFE1M	Rainfall Estimate aggregated over 1 month	☒	☒	☒
RFE3M	Rainfall Estimate aggregated over the last 3 months	☒	☐	☒
SPI1M	Standardised Precipitation Index aggregated over 1 month	☒	☒	☐
SPI3M	Standardised Precipitation Index aggregated over the last 3 months	☒	☒	☒
RCI1M	Rainfall Condition Index aggregated over 1 month	☒	☒	☒
RCI3M	Rainfall Condition Index aggregated over the last 3 months	☒	☒	☒
Month[1]	Denotes the month of the year	☐	☐	☐

[1] Variable only used in GAM models but excluded from the corresponding ANN models. The lag for the predictor variables ranges from 1 to 3 months and thus for instance, for VCI3M we consider $VCI3M_{t-1}$, $VCI3M_{t-2}$, $VCI3M_{t-3}$.

In the variable transformation step whose output is the variables in Table 13, we generated the 1-3-month lags of each of the variables for each county. The non-lagged variables were then dropped from the study except for the VCI3M that is the dependent variable. Normalization was done on all the variables to have them in the [0,1] range. Normalization is particularly useful since it transposes the input variables into the data range similar to that of the sigmoid function. Also, normalization makes all the input variables to be in a comparable range. Random sampling was used to partition the data into training and validation datasets. This follows, on the 70:30 rule for training and validation data sets, respectively.

3.7.3 Pre-study Methods

The study uses multiple indices and combines multiple methods in the prediction of drought. The prediction of drought is, for operational purposes, formulated as the prediction of future VCI3M values using the predictor variables presented in Table 13.

We focus here on predictions 1 month ahead, while 3-month ahead predictions were also tested. Two approaches are combined as shown in Figure 3.20:

- a statistical approach - Generalized Linear Models (GAM), and
- Artificial Neural Networks (ANN).

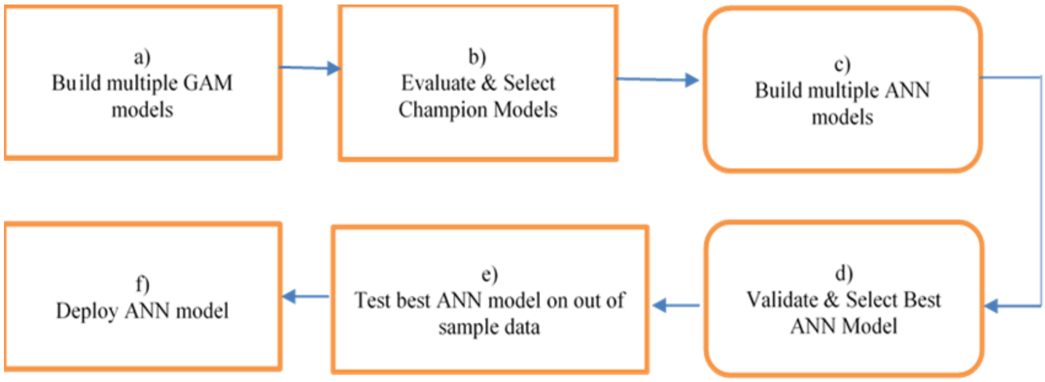


Figure 3.20: Schema of the pre-study modelling process.

As reviewed earlier and as shown in Figure 3.20 in the sub-processes (a) and (b), GAM models were used in the pre-study to arrive at the set of variables that offer the best predictions. These set of variables were then used to subsequently build ANN models.

The pre-study used the working definition of overfitting presented in Equation 30 that implies that a loss of more than 3% in performance between training and validation is deemed as overfitting.

$$Overfit\ model = \begin{cases} Yes, & diff(R2squaredT, RsquaredV) > 0.03 \\ No, & otherwise \end{cases} \tag{30}$$

Where *RsquaredT* indicates the R^2 in the training set and *RsquaredV* is the R^2 in the validation dataset

The pre-study ANNs had the configuration of 2-5-3-1 with 8 hidden nodes realized from both experimentation and the rule of thumb in Equation 31 from Huang (2003).

$$Number\ of\ nodes = \begin{cases} Sqrt\big(N * (m+2)\big) + 2 * Sqrt(N/(m+2), & 1st\ hidden\ layer \\ m * Sqrt\ (N/(m+2), & 2nd\ hidden\ layer \end{cases} \cdots \tag{31}$$

where *m* is the number of output neurons and *N* is the number of samples to be learnt with negligibly small error.

In the pre-study, the process for the execution of the artificial neural networks modelling is as presented in Figure 3.21. for the 102 models realized after the assumptions outlined earlier were enforced.

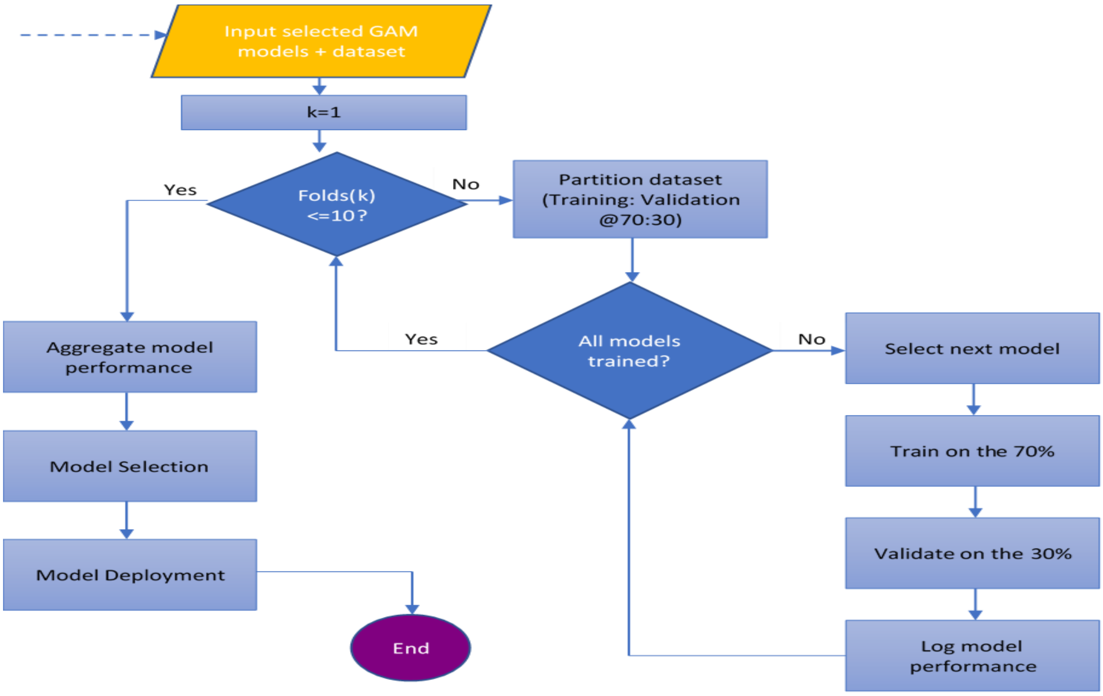


Figure 3.21. Outline of the ANN modelling process.

The ANN modelling process in Figure 3.19 sequentially inputs selected GAM models and the panel dataset followed by the iteration of the performance of the models against the data. The data is randomly partitioned in the ratio 70:30 for training and validation, respectively, for each iteration of the k times a model is run against the data. The k-fold iteration was chosen to minimize impacts of the random initialisation of the network weights.

3.7.4 Pre-study model evaluation

For all the models run, both GAM and ANN, the validation metrics used are mean absolute error (MAE), mean squared error (MSE), root mean squared error (RMSE), mean percentage error (MAPE), R^2. The results are presented using R^2. The evaluation of the performance of the models is done as part of the model training process using the validation dataset and also using the out of sample dataset.

3.7.5 GAM Model results

A plot of the performance of the 102 models in the GAM process, grouped by R^2 is presented in Figure 3.22. The models are noted to post R^2 between 0.09 and 0.86 in model training and model validation. The performances of the models in training and validation datasets (blue and orange bars respectively) indicate relative stability in model numbers across the models.

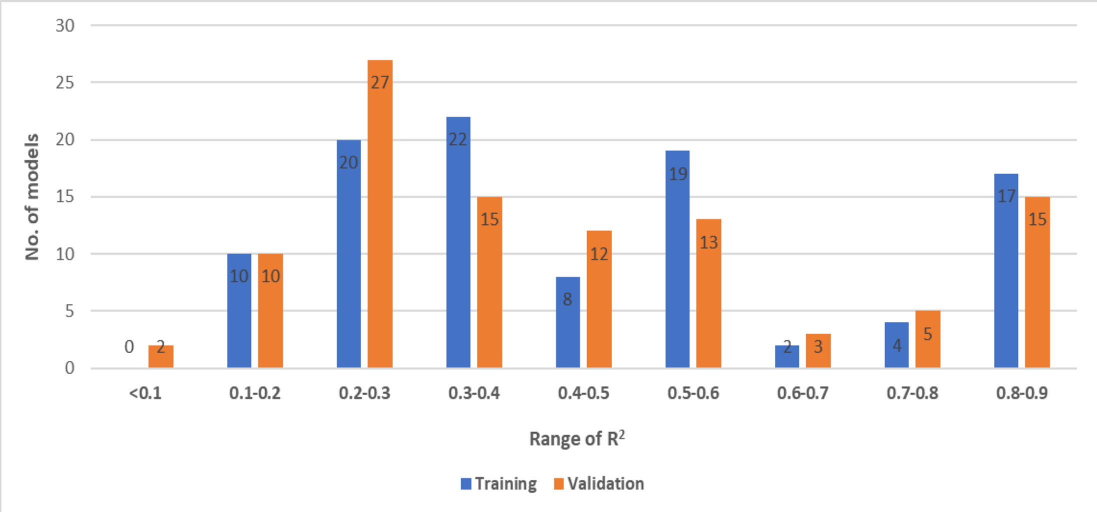


Figure 3.22: Model performance by range of R^2 in the GAM process.

The performance of the models by the lag-time of the variables, of between 1-month and 3-month lags, is provided in Figure 3.23. As expected, the analysis of the GAM process by lag time indicates that the 1-month lag of the predictors performed better in predicting VCI3M as used to define drought (in green). While a lag time of 2 months (in blue) still has some predictive power (R^2 >0.5), longer lags fail to produce good predictions (in yellow).

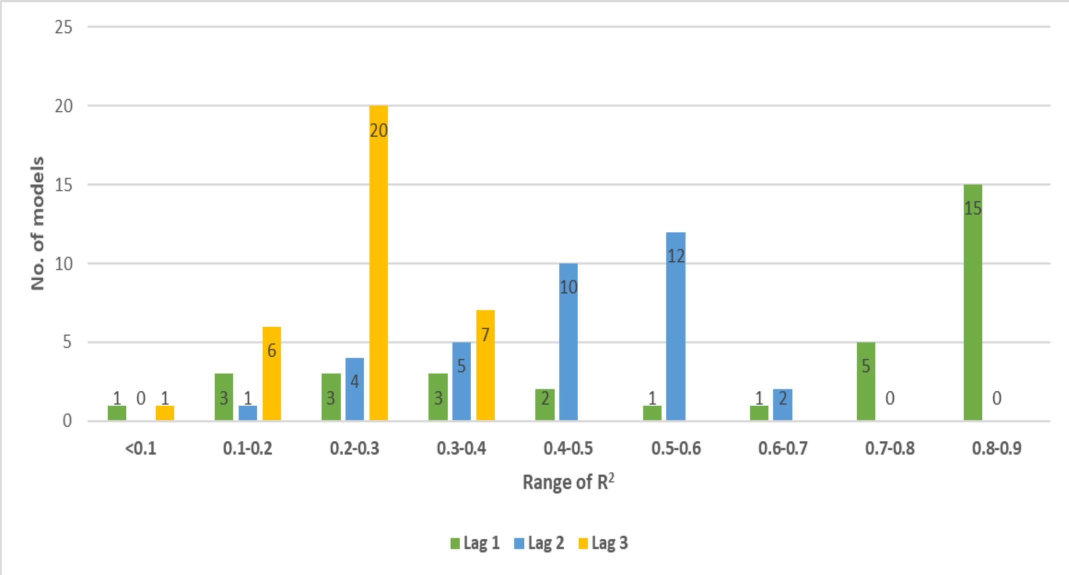


Figure 3.23: Lag-time based performance of the GAM model selection space reduction process

It is deducible that the models from the GAM process with $R^2 \geq 0.7$ as shown in Table 14 all have 1-month lag variables. In fact, the first 2-month lag variable first appears at a model ranked at position 22 with an R^2 of 0.61 while the first 3-month lag variable is in a model ranked at position 52 with an R^2 of 0.33. The poor performance of higher lags of these variables is expected since longer lags are less contributing to current vegetation status and the chances of unexpected climate variations occurring between the time of forecasting and the forecasted event increase.

Table 14: GAM models with $R^2 \geq 0.7$ in decreasing order

No	Model	R^2 Training	R^2 Validation	Overfit Index	Overfit	Lag Time
1	VCIDekad_lag1+SPI1M_lag1	0.86	0.85	0.01	No	1
2	VCIDekad_lag1+SPI3M_lag1	0.86	0.85	0.01	No	1
3	VCIDekad_lag1+RFE1M_lag1	0.85	0.85	0.01	No	1
4	VCI1M_lag1+SPI3M_lag1	0.85	0.84	0.01	No	1
5	VCI1M_lag1+SPI1M_lag1	0.85	0.84	0.01	No	1
6	VCI1M_lag1+RFE1M_lag1	0.85	0.84	0.01	No	1
7	VCIDekad_lag1+RCI1M_lag1	0.85	0.84	0.01	No	1
8	VCI1M_lag1+RCI1M_lag1	0.84	0.83	0.01	No	1
9	VCIDekad_lag1+RCI3M_lag1	0.84	0.83	0.01	No	1
10	VCIDekad_lag1+RFE3M_lag1	0.84	0.83	0.01	No	1
11	VCI1M_lag1+RCI3M_lag1	0.84	0.83	0.01	No	1
12	VCI1M_lag1+RFE3M_lag1	0.83	0.83	0.01	No	1
13	VCI3M_lag1+SPI3M_lag1	0.82	0.82	0.01	No	1
14	VCIDekad_lag1	0.81	0.80	0.01	No	1
15	VCI3M_lag1+RCI3M_lag1	0.81	0.80	0.01	No	1
16	VCI1M_lag1	0.81	0.80	0.01	No	1
17	VCI3M_lag1+SPI1M_lag1	0.81	0.79	0.01	No	1
18	VCI3M_lag1+RCI1M_lag1	0.78	0.77	0.01	No	1
19	VCI3M_lag1+RFE3M_lag1	0.78	0.77	0.01	No	1
20	VCI3M_lag1+RFE1M_lag1	0.78	0.76	0.01	No	1
21	VCI3M_lag1	0.72	0.69	0.02	No	1

The full list of all 102 models is provided in Appendix C with the full list for GAM in Table C3 and the full list of ANN models in Table C4 respectively. With the definition of overfitting in Equation 32 presented earlier, it is shown that none of the 21 GAM models with $R^2 \geq 0.7$ was judged to have suffered over-fitting. All the 21 models are thus noted to have acceptable deterioration in performance in model validation.

Though not shown, the alternative measures of performance: MAE, MSE, RMSE, MAPE etc. are noted to be consistent with R^2 since they all have a non- monotonic and non-linear relationships. An increase in R^2 translates to a change but in the reverse direction of the other error-based measures of model performance. Since the aim of the study was to use GAM modelling process as a basis for model space reduction, the above 21 models were selected for the ANN process.

3.7.6 ANN Model results

The pre-study intended to use ANNs as the case study technique of choice. Following on the model space search approach, we produced all the 21 models using the ANN process through a bagging and brute force approach in the search for the champion model. For uniformity, overfitting is defined for Artificial Neural Networks (ANN) as in GAM models.

3.7.6.1 ANN Performance in Training and Validation

Using the model overfit index in Equation 32, a few facts emerge. The ANN models were generally not overfitted as indicated in Table 15 except for only one model (No. 19) that suffered overfitting. This implies a non-overfit rate of 95%.

Table 15: ANN model performances in training and validation datasets

No	Model	Training (R^2)			Validation (R^2)			Overfit Index	Over fit
		Min	Max	Mean	Min	Max	Mea n		
1	VCIDekad_lag1+RFE1M_lag1	0.83	0.86	0.84	0.78	0.86	0.83	0.01	No
2	VCI1M_lag1+RFE1M_lag1	0.82	0.85	0.84	0.78	0.85	0.83	0.01	No
3	VCIDekad_lag1+SPI1M_lag1	0.82	0.85	0.84	0.79	0.87	0.82	0.02	No
4	VCIDekad_lag1+SPI3M_lag1	0.82	0.86	0.84	0.78	0.88	0.82	0.02	No
5	VCIDekad_lag1+RCI3M_lag1	0.82	0.86	0.84	0.79	0.87	0.82	0.02	No
6	VCI1M_lag1+SPI3M_lag1	0.81	0.85	0.84	0.78	0.87	0.82	0.02	No
7	VCI1M_lag1+RCI3M_lag1	0.82	0.85	0.84	0.79	0.86	0.82	0.02	No
8	VCI1M_lag1+SPI1M_lag1	0.82	0.85	0.84	0.77	0.86	0.82	0.02	No
9	VCIDekad_lag1+RCI1M_lag1	0.81	0.84	0.82	0.76	0.85	0.81	0.02	No
10	VCI1M_lag1+RCI1M_lag1	0.80	0.84	0.82	0.75	0.84	0.80	0.02	No
11	VCIDekad_lag1+RFE3M_lag1	0.79	0.84	0.82	0.75	0.83	0.80	0.02	No
12	VCI1M_lag1+RFE3M_lag1	0.79	0.84	0.81	0.74	0.83	0.79	0.02	No
13	VCIDekad_lag1	0.77	0.82	0.79	0.72	0.82	0.78	0.01	No
14	VCI1M_lag1	0.76	0.81	0.78	0.72	0.81	0.77	0.02	No
15	VCI3M_lag1+SPI3M_lag1	0.76	0.81	0.79	0.73	0.84	0.77	0.03	No
16	VCI3M_lag1+RFE1M_lag1	0.76	0.79	0.77	0.72	0.80	0.77	0.01	No
17	VCI3M_lag1+RCI3M_lag1	0.76	0.81	0.79	0.72	0.83	0.76	0.03	Yes
18	VCI3M_lag1+RCI1M_lag1	0.74	0.79	0.77	0.71	0.80	0.75	0.02	No
19*	VCI3M_lag1+SPI1M_lag1	0.73	0.80	0.78	0.70	0.82	0.74	0.04	Yes
20	VCI3M_lag1+RFE3M_lag1	0.71	0.77	0.74	0.65	0.76	0.72	0.02	No
21	VCI3M_lag1	0.64	0.71	0.68	0.60	0.73	0.66	0.02	No

The champion model from the ANN process is different from that of the GAM process. In fact, the ANN champion with an R^2 of 0.83 and No 1 in Table 15 was ranked the third-best model in the GAM modelling process shown earlier in Table 14 with an R^2

of 0.85. Figure 3.24 illustrates the performance of the ANN models as compared to the GAM models.

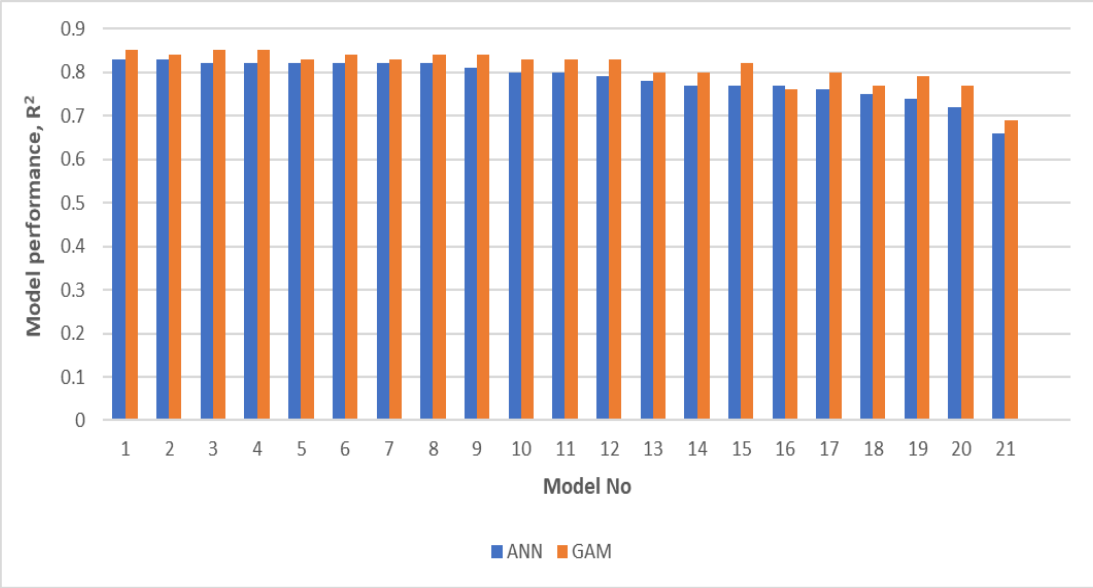


Figure 3.24: Performance of the ANN models in model validation as compared to similar GAM models

In general, as indicated in Figure 3.24, the GAM models outperform ANN models except for model 16 for which the ANN slightly out-performs GAM by an R^2 of 0.01. This is an important property since the GAM process is proved to be more optimistic in performance as compared to ANN and so fewer deserving models would be excluded from the ANN process. In training and validation, the champion model that is shown in Figure 3.25 has its best subset performance at a maximum R^2 of 0.86 has a good positive correlation between the predicted and the actual VCI3M values.

A detailed analysis of the lag-time performance of the ANN models in model training is provided in Appendix B which has the results for the 102 possible ANN models similar to those from the GAM process.

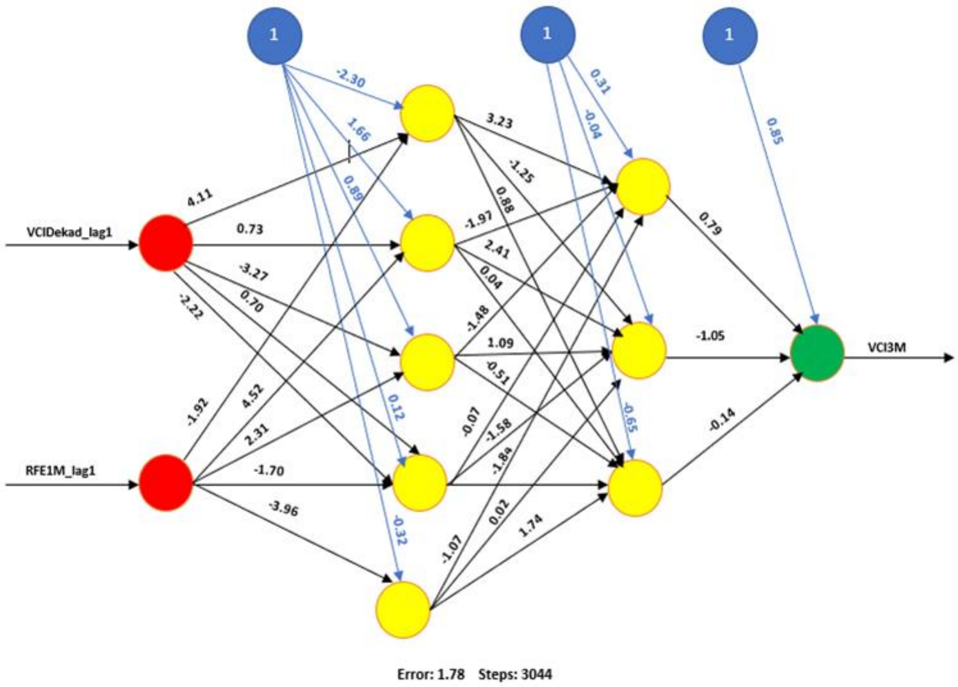


Figure 3.25: The champion NN model with the 1-month lag of the variables VCIDekad and RFE1M.

The plot of the ANN in Figure 4.25 is from the 4th partition of the training data that recorded the best performance. While blue lines indicate bias values for the ANN, the black lines represent model internal weights.

3.7.6.2 Performance of the ANN Champion in the test dataset

The out-of-sample test dataset has 96 data points across 2 years (2016-2017). The out-of-sample data was neither used in the training nor the validation processes of the ANN and even of the GAM process. It represents the model's performance in the real world and is the basis of judging the generalizability of the model.

Performance of ANN Champion* in *Regression: The ANN prediction was first formulated as a regression problem. The performance of the ANN champion in regression is indicated in the plot of the actuals versus the predicted real values for all the counties ordered by county and period as shown in Figure 3.26.

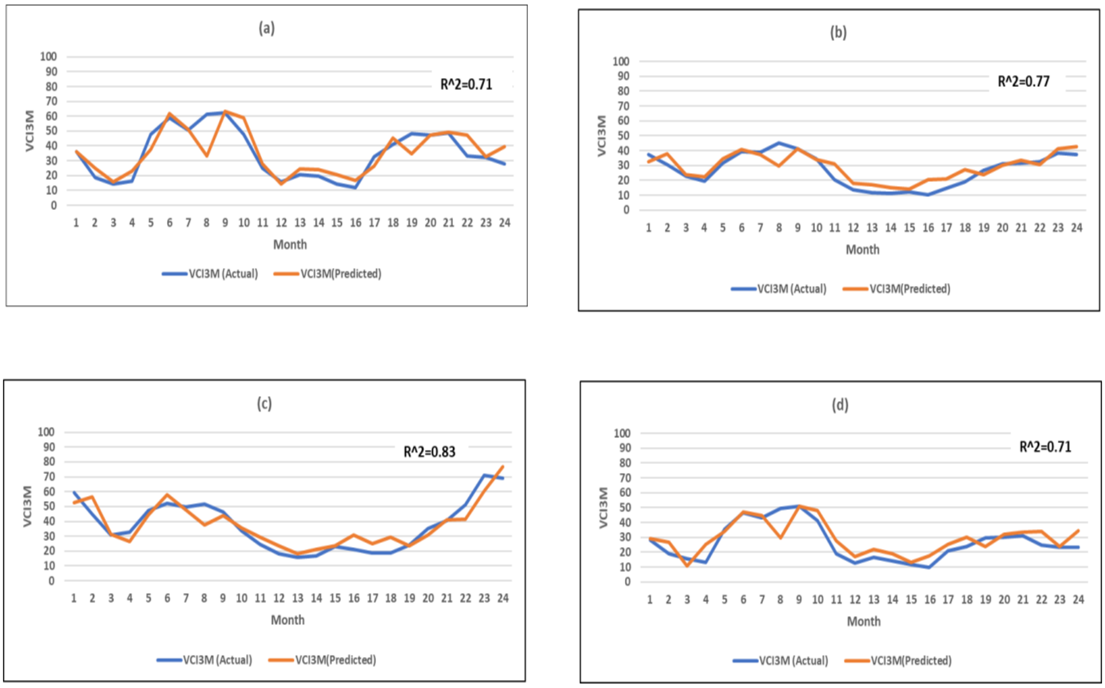


Figure 3.26: Plot of the actual versus champion model's predicted values in test data for (a) Mandera (R^2=0.71); (b) Marsabit (R^2=0.77); (c) Turkana (R^2=0.83) and (d) Wajir (R^2=0.71). Predictions were done 1 month ahead.

The plot of the actual versus the predicted values represents quite a good agreement. In the test data, the champion model posted an R^2 of 0.78 and unscaled RMSE of 7.03 on the actual data values. The above performance over the 96 data points for testing is acceptable in the prediction of future drought events.

Performance of ANN in classification: Operational drought monitoring involves the definition of thresholds on indices used for drought monitoring to realise a class approach to drought monitoring. We use the approach documented in Klisch & Atzberger (2016), Meroni et al. (2019) and Klisch, Atzberger & Luminari (2015) as presented in Table 16 to monitor drought in five drought classes:

Table 16: Classification of drought based on vegetation deficit classes

VCI3M Limit Lower	VCI3M Limit Upper	Description of Class	Drought class
≤0	<10	Extreme vegetation deficit	1
10	<20	Severe vegetation deficit	2
20	<35	Moderate vegetation deficit	3
35	<50	Normal vegetation conditions	4
50	≥100	Above normal vegetation conditions	5

The champion model had an overall accuracy of 67% rising to 71% for Wajir and Marsabit counties as indicated in the matrix provided in Figure 3.27.

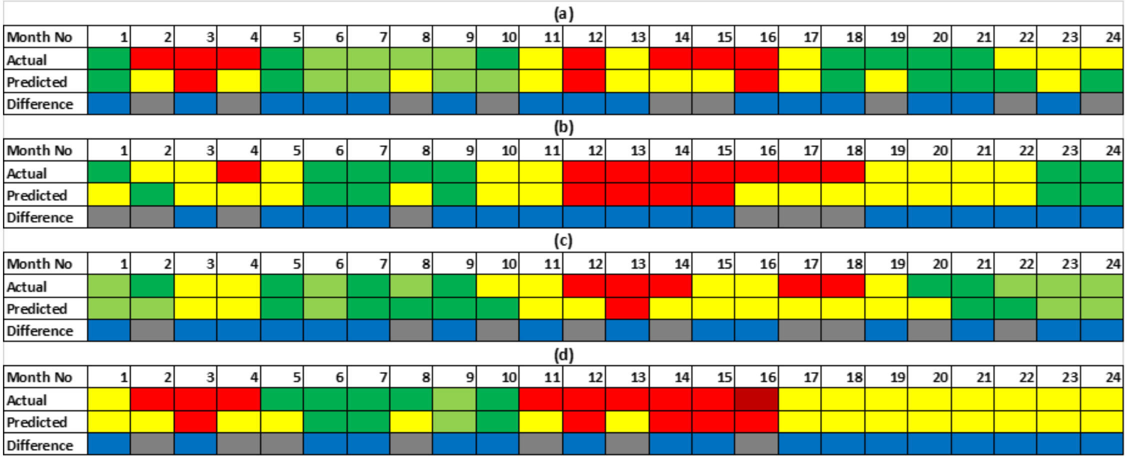


Figure 3.27: Performance of the ANN classifier for each of the counties

The months of difference (Figure 3.27) are shown in grey and those of agreement in blue. Predictions are done 1 month ahead. The classification accuracies are: **(a)** 63% for Mandera county; **(b)** 71% for Marsabit county; **(c)** 63% for Turkana county and; **(d)** 71% for Wajir county.

When formulated as a multi-class classification problem and the multiple receiver operating (ROC) curves plotted for each of the pairwise comparisons following the approach in Hand & Till (2001), we obtained the ROC curve plotted in Figure 3.28.

The multi-class area under the curve (AUROC) is the average of the 10 areas under all the ROCs. The ROC for the 5 classes provides a reasonable trade-off between sensitivity and specificity at an overall AUROC of 89.99%. The area under the ROC (AUROC) indicates quite a good trade-off between sensitivity and specificity and is ranked within the good performance category as it is way above the 50% that represents a worthless test (in grey).

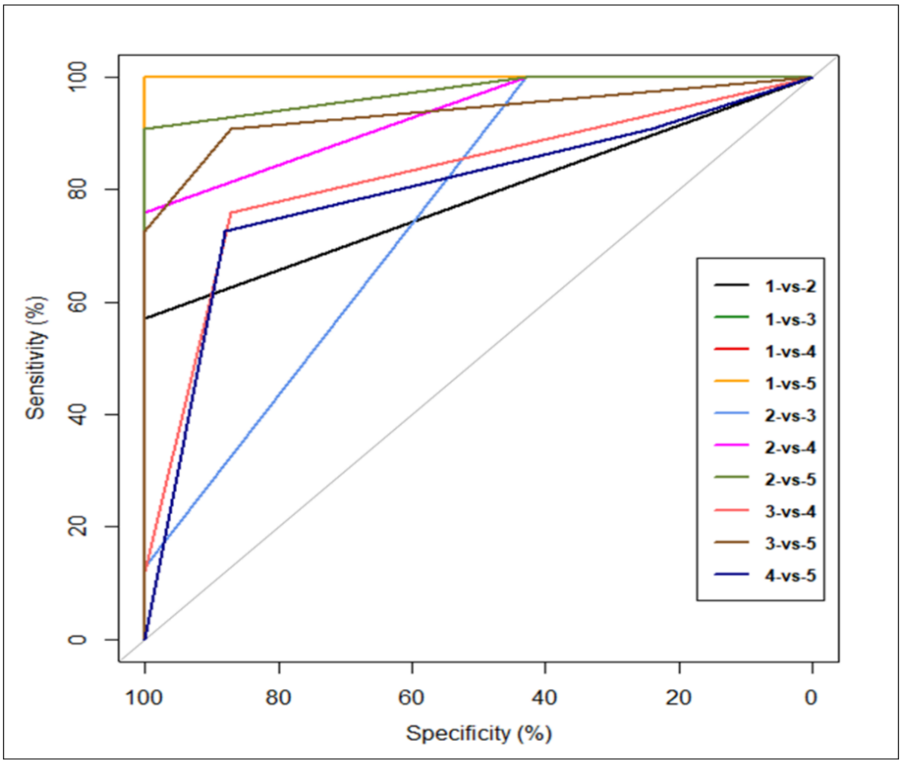


Figure 3.28: Multi-class ROC plot of the champion model as a drought phase classifier.

The curves in the multi-class ROC plot (Figure 3.28) represent the pairwise comparison of the five (5) classes. The overall area under the multi-class ROC is the average of the areas under each of the ROCs for the pairwise class comparisons.

3.7.7 Validation of the key assumption of the pre-study

3.7.7.1 Appropriateness of GAM technique in variable selection

To validate the key assumption on the appropriateness of the GAM modelling technique in the reduction of the model space, we run the extra 81 models through the ANN process. The best performer from the set of non-selected models had an R^2 of 0.50. A summary of the performance of the non-GAM selected models in the test dataset is provided in Figure 3.29.

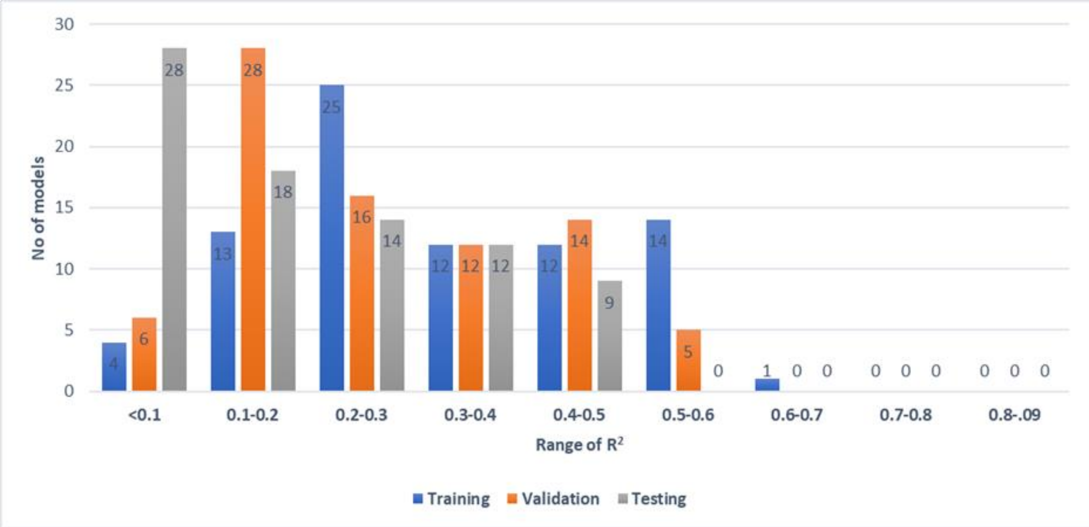


Figure 3.29: Distribution of non-selected models ANN performance on training, validation and testing.

The results in Figure 3.29 indicate no model posted an R^2 of at least 0.5 in model testing. The results validate the assumption of the utility of GAM in modelling non-linearity as well as in their use in this study for model space reduction prior to the use of computationally intensive algorithms like artificial neural networks. The models that were not selected by the GAM process are not expected to perform any better in the ANN process than the GAM selected models. The GAM process is, in essence, more optimistic in performance ranking than the ANN process. This property is useful as it generally guarantees that good models are not filtered out of the ANN process.

3.7.7.2 Investigation of multi-collinearity

The collinearity-matrix in Figure 3.30 gives the correlation coefficients between the predictor (X) variable pairs together with a proposed interpretation scheme. The absolute correlation coefficient between the pairs in X is provided together with a proposed interpretation of the correlations.

	VCI3M_lag1	NDVIDekad_lag1	VCI1M_lag1	VCIDekad_lag1	RCI1M_lag1	RCI3M_lag1	RFE1M_lag1	RFE3M_lag1	SPI1M_lag1	SPI3M_lag1
VCI3M_lag1	1.00	0.53	0.86	0.87	0.06	0.41	0.00	0.26	0.10	0.45
NDVIDekad_lag1	0.53	1.00	0.55	0.55	0.16	0.30	0.19	0.49	0.04	0.31
VCI1M_lag1	0.86	0.55	1.00	1.00	0.18	0.52	0.09	0.31	0.18	0.54
VCIDekad_lag1	0.87	0.55	1.00	1.00	0.18	0.53	0.09	0.31	0.17	0.54
RCI1M_lag1	0.06	0.16	0.18	0.18	1.00	0.51	0.86	0.49	0.60	0.42
RCI3M_lag1	0.41	0.30	0.52	0.53	0.51	1.00	0.38	0.67	0.46	0.86
RFE1M_lag1	0.00	0.19	0.09	0.09	0.86	0.38	1.00	0.57	0.31	0.29
RFE3M_lag1	0.26	0.49	0.31	0.31	0.49	0.67	0.57	1.00	0.14	0.48
SPI1M_lag1	0.10	0.04	0.18	0.17	0.60	0.46	0.31	0.14	1.00	0.54
SPI3M_lag1	0.45	0.31	0.54	0.54	0.42	0.86	0.29	0.48	0.54	1.00

Interpretation of correlation	
	=0.0: No linear relationship
	0.0 - <0.3: Weak
	0.3 - <0.5: Low
	0.5 - <0.7: Moderate
	>=0.7: High to vey high

Figure 3.30: Collinearity-matrix for the input (X) variables.

From the collinearity matrix in Figure 3.30, the correlation between vegetation input variables is between moderate to very high correlations (min=0.53, max=0.87). This is as opposed to say the relationship between the pairings between vegetation and precipitation datasets that is between no linear relationship to moderate (min=0.0, max=0.54). The assumption to avoid the use of datasets of the same type whether precipitation or vegetation datasets and in essence use the pairings between precipitation and vegetation datasets generally results in pairings of weak to barely moderate correlations. This is the first step to avoiding the problem of multi-collinearity.

In addition to the collinearity matrix, we investigated the problem of multi-collinearity between the independent variables in a two-step process – first for a model of all variables and second for the pairing of precipitation and vegetation variables. For each approach, we obtained the variance inflation factor (VIF) with the rule of thumb that a VIF>5 indicates high multi-collinearity while a VIF>10 indicates multi-collinearity that has to be handled in the modelling process. The thresholds for the VIF are generally a rule of thumb and are, for example, discussed in Mathieson, Peacock & Chin (2001) and in Kock & Lynn (2012). The results of the investigation of VIF for all the model variables are presented in Table 17.

Table 17: Variance inflation factor (VIF) for a single model with all 1-month lag variables

Variable	Variance inflation factor (VIF)
VCI3M_lag1	6.14
NDVIDekad_lag1	1.41
VCI1M_lag1	976.21
VCIDekad_lag1	1,057.46
RCI1M_lag1	4.41
RCI3M_lag1	5.90
RFE1M_lag1	2.63
RFE3M_lag1	2.88
SPI1M_lag1	3.34
SPI3M_lag1	5.24

The full model with all variables indicates the presence of multi-collinearity with VIF>10 for 2 of the predictor variables. Further analysis for the models fed into the GAM process obtains the results provided in Figure 3.31.

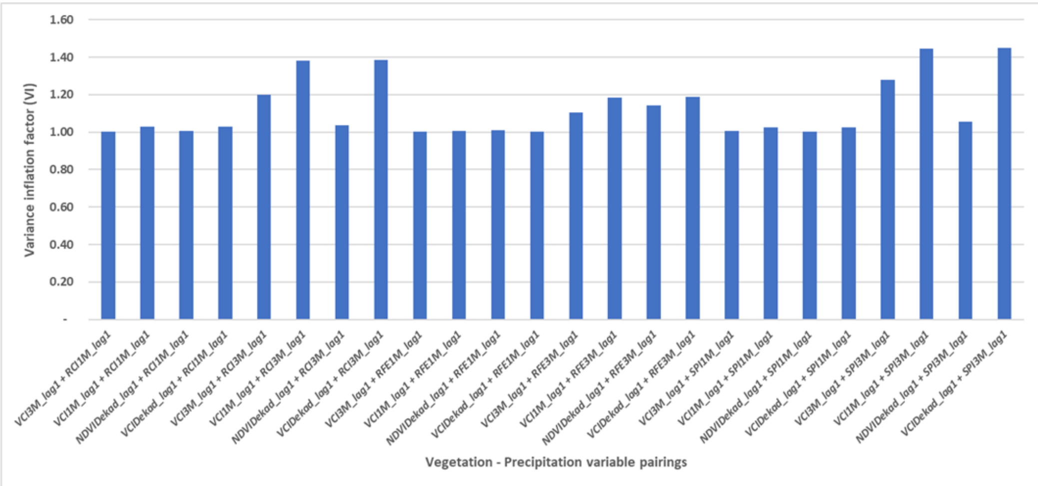


Figure 3.31: Variance inflation factor (VIF) for GAM models

The results in Figure 3.31 confirm that using the vegetation-precipitation variable pairs that correspond to the low correlation portions of the correlation matrix in Figure 14 ensures models that are not affected by multi-collinearity.

Concurvity, that has similar effects to those posed by multi-collinearity, was not a major limitation in the approach to GAM modelling employed by this study since only one smooth term was used in the development of the GAMs across all the models. Additional investigation with smoothing on all the terms, though not presented, resulted in massive model overfitting that then limited the smoothing of the predictor variables.

3.7.7.3 Most appropriate lag time

The investigation of the most appropriate lag time in the prediction of future droughts is presented in Appendix B. The lag-time performance of both the GAM (Figure B1 and Table B1) and the ANN (Figure B2 and Table B2) indicate a few interesting facts:

- The models considered predictive ($R^2 \geq 0.7$) were only posted when predictions were made 1 month ahead. In fact, very good performance at a maximum of $R^2=0.83$ for GAM and $R^2=0.83$ for ANN is posted in predictions of conditions 1 month ahead.
- The predictions 3-month ahead post very weak models whose performance are at best $R^2=0.33$ for GAM and $R^2=0.25$ for ANN. These models are considered to perform at below chance.
- It is evident that the only viable predictions are from models that make predictions 1 month ahead.

3.7.7.4 Performance of models with multiple lags of the same variable

Non-significance of gain in performance from multiple lags of the same variable was one of the assumptions of the study that was tested in the pre-modelling. This was a two-step process. First, we tested for the possibility of multi-collinearity when multiple lags of the same variable are used in the same model. Only three models of the 40 returned a variance inflation factor, VIF >5. Having multiple lags of the same variable do not, therefore, suffer multicollinearity. Despite these results indicating that multiple lags of the same variable can be used in the same model for the balance of the 37 models, an investigation of their performance indicates the contrary as documented in Adede et al. (2019a). Only 17 models post a performance gain of 1 percentage point or more. Furthermore, all the models that post an R^2 of at least 0.5 either have a loss of 6% in performance to a gain of 1% in performance implying poor returns in having multiple lags of the same variable in a model.

3.7.8 Summary of the pre-study

In this pre-study, multiple variables were used to predict future vegetation condition index aggregated 3-monthly as a proxy to drought conditions. The predictor variables were 1-3-month lags of precipitation and vegetation indices. The methodology used two techniques in a hybrid setup where the General Additive Model (GAM) statistical approach was first run against several possible model configurations. The GAM method was then used to reduce the model space and by extension the set of viable variables. After variable selection and with the model space reduced, a brute force approach was then employed using the Artificial Neural Networks (ANN) approach.

One month ahead forecasts of the VCI using the best ANN model showed good performances with R^2 ranging between 0.71 and 0.83. After grouping into five drought classes, 63% to 71% of the months were correctly classified across the counties with the remaining months showed a maximum difference of one class. Prediction skills deteriorated with lag times longer than 1 month. The poor performance of variables with longer times lags, in the prediction of drought events was established. Since the approach builds multiple models before evaluation in the search for the best model, it is possible to support model ensembling that supports the use of multiple models in the prediction of future events.

The study demonstrates that model space reduction is beneficial to the building of neural networks that are known to generally have slower training times as compared to other approaches. The automation of the model training and model validation processes, and the measure of performance with a view to quantifying and avoiding overfitting, make for a scalable approach.

3.7.9 Criticisms of the Pre-study

The pre-study realised relatively predictive models that had a relatively good performance. The model space search approach was helped by the use of GAMs that were faster to compute. The performance of the champion ANN model however suffered poor performance in the prediction of moderate to extreme drought that are, in fact, the classes for which best performance is desired. The problem of performance in the moderate to extreme drought classes is illustrated in Figure 3.32.

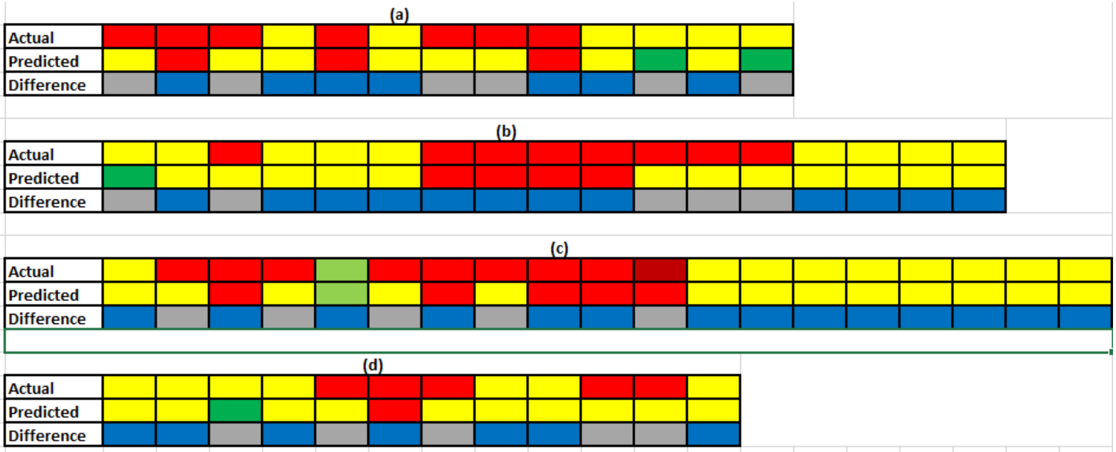


Figure 3.32: Performance of the ANN champion in moderate to extreme drought.

Figure 3.32 shows the performance of the champion ANN in the prediction of moderate to extreme droughts for each of the counties. The months of difference are shown in grey while those of agreement are shown in blue. Predictions were done 1 month ahead. The classification accuracies recorded were: (a) 54% for Mandera county; (b) 71% for Marsabit county; (c) 58% for Turkana county and; (d) 74% for Wajir county.

Further analysis for severe to extreme drought, however, returned very poor performance perhaps due to low occurrence of the events in the training data at 4.92% and 10.81% for severe and extreme droughts respectively.

One possible mitigation to this poor performance in class distribution would be model ensembling. Given that the ANN process realized 21 models that were relatively good performers, we can have all the models participate in the prediction process. A naive approach to model ensembling would be to average the scores from all the models prior to the classification. This approach realizes an overall R^2 of 0.81 and an overall accuracy of 74%. At the county level, the performance was: Mandera (R^2=0.70), Marsabit (R^2=0.82), Turkana (R^2=0.87) and Wajir (R^2=0.76). The performance of the classification by county for moderate to extreme drought is provided in Figure 3.33.

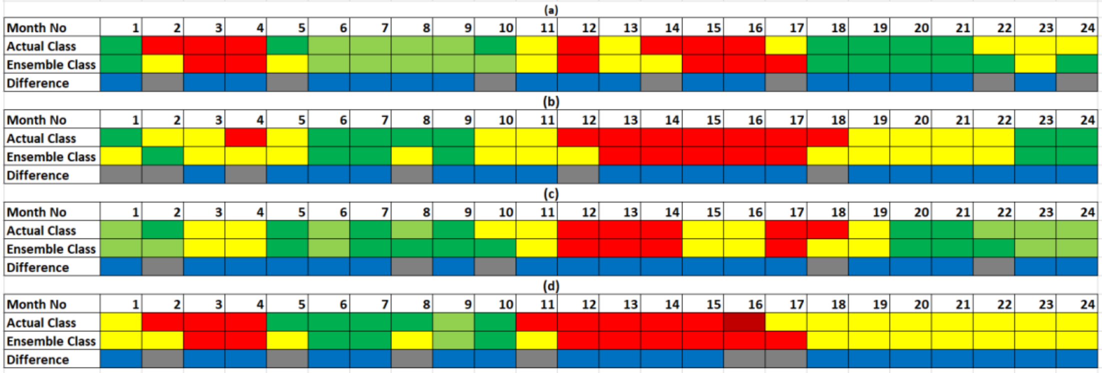


Figure 3.33: Performance of the average ensemble classifier for all the vegetation deficit classes.

For each of the counties in Figure 3.33: Mandera, Marsabit, Turkana and Wajir respectively, the months of difference are shown in grey while those of agreement are in blue. Predictions are done 1 month ahead. The classification accuracies in the severe to extreme vegetation classes are: (a) 71% for Mandera county; (b) 63% for Marsabit county; (c) 80% for Turkana county and; (d) 67% for Wajir county.

There is a gain in classification accuracy that is realized when the model ensembling approach is used as compared to the use of the single best model. The model ensembling approach, however, has a computational resource and time aspect to it that adds to the limitations of the ANNs that already suffer the limitation of interpretability of both process and output.

The main study investigated the question of the use of model ensembling to realise more predictive models in detail using different model ensembling techniques. The results of this study are documented in Chapter 4- Results and Discussion.

Chapter 4: RESULTS AND DISCUSSION

In this section, we present the results from the study based on the correlation research design that was deployed as described in the previous section. The research followed the aspects of exploratory data analysis (EDA) sequenced with predictive modelling on the variables selected from the EDA process. We document the results following on the research objectives and for each objective, we logically formulate the questions that guide the study to provide the assurance that the objectives are met.

4.1 Variables used for drought monitoring

To meet the study objective of "*Determine the different biophysical and socio-economic variables that are used in the monitoring/ prediction of drought and investigate their relationship with drought.*", we formulated two research questions that we use to organise the results. These questions, used to determine the order the results are presented, are as follows: -

- **RQ1:** What are the different biophysical and socio-economic variables that are used in the monitoring/ prediction of drought? This is handled in sections 4.1.1 that tabulates these variables as identified from literature and 4.1.2 that identifies the key relationships between these variables.
- **RQ2:** How do the variables identified for drought monitoring relate with drought and drought effects? This is handled in section 4.1.3

4.1.1 Identification of the variables used in the monitoring/ prediction of drought

To answer the first research question (RQ1), we surveyed literature intending to document variables used in both drought monitoring and the prediction of drought. We also sought to have the entire spectrum of variables used for this purpose. Answering this question involved two key undertakings: *(1)* a search through literature on past studies to identify the variables documented for use in both drought monitoring including their respective data archives; *(2)* the download and pre-processing of the datasets identified to realize the variables for the area of study covering the period of study.

The four drought types from literature review and theology formed the basis of the identification of the variables used for drought monitoring. The base datasets that were: NDVI, LST, EVT, PET, SPEI, MUAC, Maize Price & Goat Price were then transformed to realize the indicators whose summary statistics are provided in Table 18.

Table 18: Summary statistics of the indicators processed for drought monitoring

No	Drought Type	Variable	Min	Max	Range	Mean	Median	Stdev
1	Met	TAMSAT_RFE1M**	0	191.10	191.10	29.29	14.20	35.07
2		CHIRPS_RFE1M**	1.04	236.68	235.64	26.91	13.20	36.16
3		TAMSAT_RFE3M***	0	114.23	114.23	29.29	26.47	22.01
4		CHIRPS_RFE3M***	2.46	131.22	128.76	27.23	21.95	21.89
5		TAMSAT_SPI1M**	-1.87	2.37	4.24	0.25	0.26	0.67
6		CHIRPS_SPI1M**	-2.36	4.27	6.63	0.05	0.25	0.80
7		TAMSAT_SPI3M***	-2.24	2.55	4.79	0.11	0.12	0.75
8		CHIRPS_SPI3M***	-2.69	3.22	5.91	0.02	-0.07	0.86
9		TAMSAT_RCI1M**	0	98.85	98.85	24.48	19.10	22.41
10		CHIRPS_RCI1M**	0	99.16	99.16	21.87	15.28	19.37
11		TAMSAT_RCI3M***	0.01	99.78	99.77	31.69	28.39	20.09
12		CHIRPS_RCI3M***	0	99.14	99.14	29.73	24.24	21.81
13	Agric	NDVIDekad*	0.15	2.69	2.54	0.43	0.25	0.54
14		VCIDekad*	3.81	99.17	95.36	36.17	32.68	19.51
15		VCI1M**	4.20	99.09	94.89	36.53	32.81	19.25
16		VCI3M***	5.43	87.06	81.63	35.69	33.67	16.92
17	Hydro	LST1M**	298.51	311.45	12.94	305.78	305.98	2.42
18		TCI1M**	0.75	137.02	136.27	57.22	58.18	21.59
19		EVT1M**	1.65	56.75	55.10	10.04	7.85	7.57
20		PET1M**	48.36	303.08	254.72	150.38	152.02	50.35
21		SPEI1M**	-4.32	2.21	6.53	-0.45	-0.40	1.10
22		SPECI3M***	-4.14	1.88	6.02	-0.51	-0.42	1.11
23	SED	TOT**	13.87	113.44	99.57	45.74	42.21	19.02
24		MUAC**	10.80	36.00	25.20	22.05	21.40	5.06

Note: Met, Agric, Hydro and SED refer to Meteorological, Agricultural, Hydrological and Socio-economic drought respectively.

From the initial set of variables, two of the socio-economic data of Maize Price and Goat Price were converted to Terms of Trade (ToT) and the initial variables dropped from the model as earlier indicated in methodology. Together with the results presented later on the selection between the TAMSAT and CHIRPS datasets, the final variables were: Meteorological (6), Hydrological (6), Agricultural (4) and Socio-Economic (2).

The range of values in Table 19 for the raw data shows very diverse ranges making the values quite difficult to model on. The need to have the values in comparable ranges supports the need for normalization whose results are presented later.

4.1.2 Descriptive analysis of the identified data for drought monitoring

The exploratory investigation into understanding the variables to be used in the study and the relationships between them was done as part of the pre-study. In this section, we present the characteristics of the data collected using both descriptive analysis processes and also the relationships between them. Section 4.1.3 handles the relationships between the variables and both drought severity and drought effects.

4.1.2.1 Characteristics of the datasets

The summary statistics for each of the datasets is presented following on their groups in the study of drought.

Characteristics of the Precipitation data

Table 19 presents the descriptive statistics for the precipitation datasets (TAMSAT and CHIRPS) used majorly in the monitoring of meteorological droughts. The non-transformed data (RFE1M) is the basis of the analysis of the differences for the entire region.

Table 19: Descriptive statistics for monthly rainfall estimates (RFE) for each county in the study area

County	Dataset	Skewness	Excess Kurtosis	Mean	Median	StDev
Mandera						
	TAMSAT_RFE1M	1.83	2.72	23.28	5.45	35.16
	CHIRPS_RFE1M	2.24	4.98	27.31	5.77	41.51
Marsabit						
	TAMSAT_RFE1M	1.63	2.52	27.54	13.49	31.96
	CHIRPS_RFE1M	2.39	6.49	25.41	11.91	33.99
Turkana						
	TAMSAT_RFE1M	1.02	0.50	40.06	31.47	33.36
	CHIRPS_RFE1M	1.63	2.56	27.35	20.20	22.44
Wajir						
	TAMSAT_RFE1M	1.80	2.87	26.29	8.46	37.14
	CHIRPS_RFE1M	2.40	6.11	27.58	5.82	42.88

The precipitation datasets, analysed at county level indicated competitive average monthly rainfall for the period across the counties and both TAMSAT and CHIRPS except for Turkana. Turkana gives the widest difference in the monthly average rainfall at 12.71mm for the period between TAMSAT and CHIRPS. The annual average rainfall from the modelled approach realised higher approximations for Mandera, Marsabit and Turkana while Wajir had the annual rainfall under approximated. An example of a competitive approximation is the rainfall for Mandera county for the period that translates to an average rainfall of between 279mm and 328mm. This compares favourably with the annual averages for Mandera of around 250mm. The deviation could be attributed to the tendency for the cold cloud duration models to over-estimate rainfall due to cases when cold clouds do not lead to actual rainfall. Clearly, for all the counties, monthly average rainfall from both TAMSAT and CHIRPS have general concurrence except for Turkana where there is a major divergence. The same divergence is recorded for Turkana based on the median values.

From Table 19, the TAMSAT and CHIRPS rainfall estimates are positively skewed and are hence skewed to the right. Using the rule of thumb (Bulmer, 1979) to interpret the skewness numbers, the datasets are considered highly skewed. This is because all the datasets reported skewness greater +1 that implies highly skewed to the right.

The reported kurtosis values in Table 19 are referred to as excess kurtosis and equals kurtosis over or below 3 that denotes the kurtosis for a normal distribution. The two datasets have an excess kurtosis greater than zero across all the counties. The distribution of the datasets is thus referred to as **leptokurtic**. This implies that the precipitation datasets do not conform to normality. A comprehensive comparison with a view to dataset selection, and hence variable selection, for the precipitation datasets is presented in section 4.1.3.

Characteristics of the Vegetation data

The vegetation datasets are derived from NDVI as the base dataset. The NDVI dataset was then processed to vegetation condition index (VCI) at 1-monthly and 3-monthly aggregation timelines. To realize the NDVI data for the study in an objective process, we evaluated three NDVI datasets. The summary results for the descriptive analysis of the three datasets for the period 2001-2015 is provided in Table 20. The three datasets were disaggregated at the county level.

Table 20: Descriptive statistics for three NDVI datasets (Boku, Own, FewsNet)

Dataset	Mean	Median	Variance	Kurtosis	Skewness	Range
Turkana						
NDVI_Boku	0.25	0.23	0.00	0.04	0.83	0.25
NDVI_Own	0.25	0.24	0.00	(0.21)	0.73	0.22
NDVI_FewsNet	0.24	0.23	0.00	0.18	0.80	0.25
Mandera						
NDVI_Boku	0.31	0.27	0.01	0.88	1.26	0.45
NDVI_Own	0.31	0.27	0.01	1.15	1.19	0.41
NDVI_FewsNet	0.29	0.27	0.01	2.12	1.35	0.44
Marsabit						
NDVI_Boku	0.23	0.21	0.00	2.11	1.47	0.35
NDVI_Own	0.23	0.21	0.00	2.06	1.43	0.29
NDVI_FewsNet	0.22	0.20	0.00	3.80	1.64	0.37
Wajir						
NDVI_Boku	0.28	0.24	0.01	4.95	1.92	0.52
NDVI_Own	0.28	0.25	0.01	4.64	1.87	0.44
NDVI_FewsNet	0.19	0.16	0.01	5.23	1.97	0.44

The three datasets have different pre-processing methods but the same underlying data source. It is evident from the results in Table 20 that pre-processing does not alter the data characteristics in any major ways since the descriptive statistics realize very close results across the counties. This implies that for drought monitoring, consistency in the processing steps could be more important than complexity in the processing chains adopted as long as key processing steps are fully carried out on the base data. This is particularly so given that the BOKU dataset, in this case, represents state of the art pre-processing using Whittaker weighted smoothing and uses both Aqua and Terra satellites (Klisch & Atzberger, 2016). This BOKU processing is as compared to the FEWSNET dataset (FEWSNET Data Portal, 2016) that uses solely the Terra MODIS satellite and is smoothed using the Swets algorithm (Swets et al., 1999). Further, even the OWN dataset that uses a Naïve implementation of the Whittaker smoothing algorithm and used only the Terra sensor data seems competitive to the other datasets.

The test for normality for the vegetation datasets for randomly selected pixels across the images returned the histogram plot provided in Figure 4.1 that confirms the non-normality in the distribution of the base datasets even at the image level.

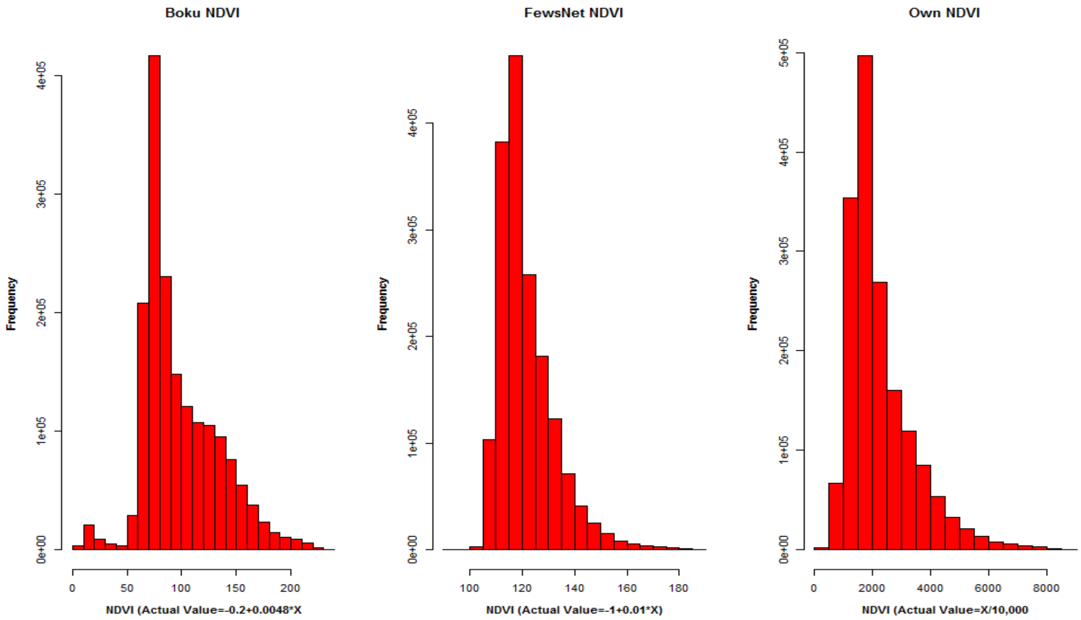


Figure 4.1: Histogram plots for cell level values of Boku, FewNet and Own datasets respectively

A correlational analysis of the data following on spearman's correlation coefficient returned a correlation coefficient of 0.9 between Boku and FewsNet datasets; 0.92 between Boku and Own datasets and 0.98 between Own and Fewsnet datasets. This is because of the similarity between the Fewsnet and the Own datasets whose differences is majorly attributed to the approach used in smoothing.

A time series analysis on the three datasets was done using the additive time series model in Equation 32. The aim was to break down each time series into its principal components including a seasonal component -S_t, a trend component-T_t and a remainder component-E_t.

$$y_t = S_t + T_t + E_t \quad \text{..........(32)}$$

A plot of the trend components for the decomposed vegetation datasets is presented in Figure 4.2.

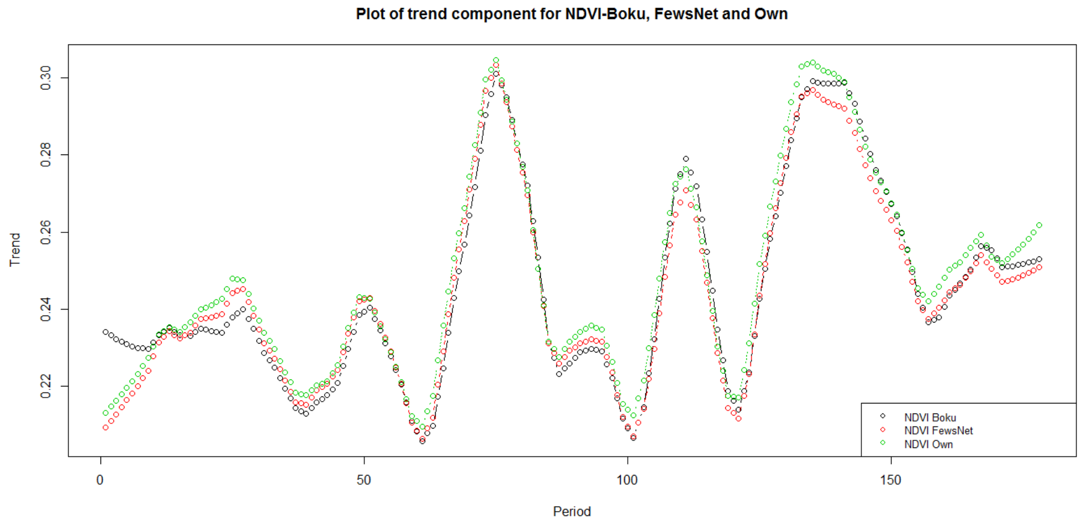


Figure 4.2: Plot of the trend component from the decomposition of the vegetation datasets

The similarity in trend for the three datasets is confirmed by the high correlation coefficients provided in Table 21.

Table 21: Correlation coefficient matrix for trend component between the pairs of vegetation datasets

Dataset	Spearman's Correlation Coefficient
Boku - FewsNet	0.97
Boku - Own	0.97
Fewsnet - Own	1.00

Adjusting for seasonality and leaving the trends and error components in the datasets intact gives the plot in Figure 4.3. and the correlations presented in Table 22.

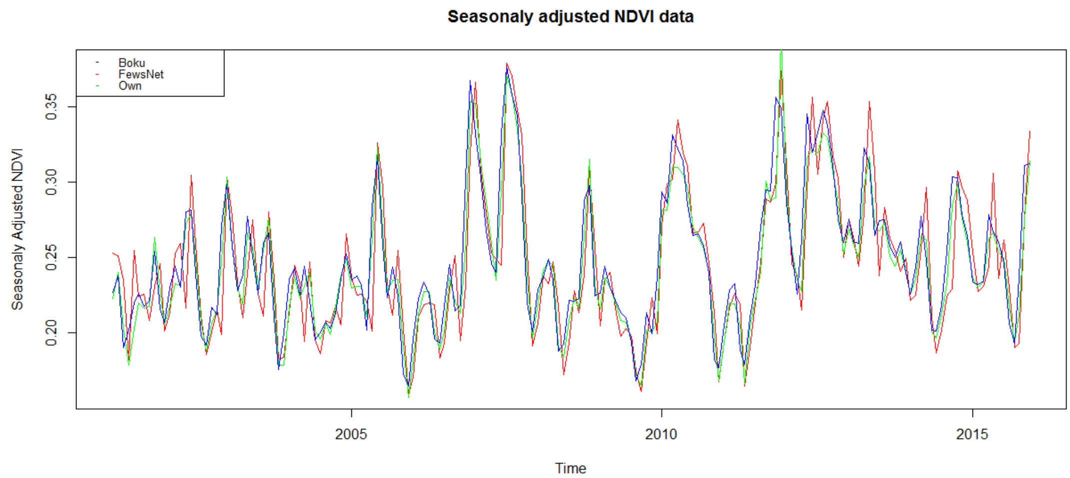


Figure 4.3: An integrated comparison of the three seasonally adjusted vegetation datasets with Boku NDVI in blue, FewsNet NDVI in red and Own NDVI in green.

Table 22: Correlation coefficient matrix for seasonally adjusted NDVI datasets

Dataset Pair	Spearman's Correlation Coefficient
Boku - FewsNet	0.891
Boku - Own	0.805
Fewsnet - Own	0.962

The visual inspection amongst the three seasonally adjusted vegetation datasets in Figure 4.3 shows some good agreement which is confirmed by the correlations matrix in Table 22. The seasonally adjusted datasets have a minimum correlation of 0.805 between the Boku and Own datasets and a maximum of 0.962 between the Fewsnet and Own pairs.

A visualization of the month on month plot of the seasonal variation components of the three datasets provided in Figure 4.4 illustrates the occurrence of two distinct peak seasons of greenness between the months of May – June and November-December across the entire years of the three vegetation datasets.

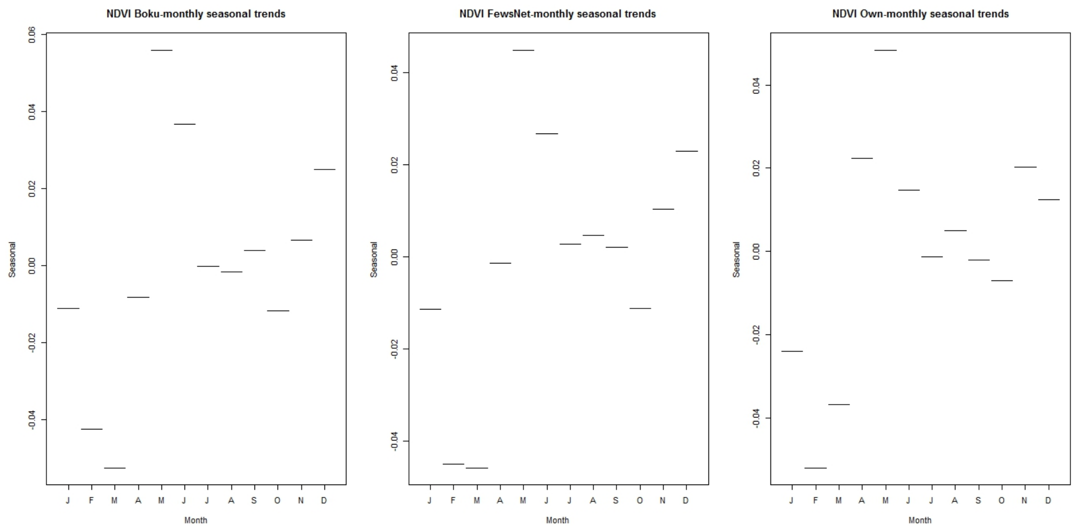


Figure 4.4: Plot of the seasonal component from the decomposition of the vegetation datasets

The rise in vegetation cover as shown in Figure 4.4 for the three datasets occurs in two distinct phases of from March through April to May period (MAM) and October through November to December (OND). These periods corresponded to the long and short rainy seasons in Kenya.

Relationship between precipitation and vegetation datasets

The final investigation of the relationship between the variables for the prediction of drought severity was the investigation of the relationship between the vegetation and precipitation datasets. An investigation of the maximum lag between the precipitation and vegetation datasets for each county produced the results provided in Table 23.

Table 23: Lag correlation between precipitation and vegetation

County	CHIRPS Correlation	TAMSAT Correlation	CHIRPS Lag	TAMSAT Lag
Turkana	0.71	0.70	-1	-1
Mandera	0.71	0.67	0	0
Marsabit	0.76	0.73	-1	-1
Wajir	0.71	0.73	-1	-1

The results of the determination of maximum cross-correlation between the dataset pairs by county suggest that the response of vegetation to rainfall is more pronounced in Mandera County that has a shorter lag of below 1 month. The shorter time lag for Mandera implies that vegetation responds faster to changes in precipitation in Mandera as compared to the other counties. Vegetation as having a 1 month lag in response to changes in rainfall remains an expected scenario since the study uses both vegetation and precipitation datasets that are aggregated monthly. A finer time step would be needed for a finer analysis to quantify the vegetation and precipitation lag in the order of weeks.

A summary into the investigation of the vegetation datasets is that both the precipitation data and vegetation datasets the underlying datasets for vegetation are not normally distributed. This is even observed in the transformed data from each of the base sources. Any modelling on the data needed to take care of this fact and non-parametric approaches were settled on as the basis for further data analysis. It is equally the case that there exist close agreements between the datasets and therefore the study's process to pre-processing remote sensing data was adopted for the study.

Mapping results to RQ1 to answer the research question " *What are the different biophysical and socio-economic variables that are used in the monitoring/ prediction* *of drought*?" was thus achieved by tabulating the variables earlier surveyed in literature and presenting the data used in the generation of historical data for these variables. The variables are presented based on what aspects of drought they are expected to quantify. The basic descriptive statistic of the non-normalized and non-lagged data was undertaken. For the multiple vegetation datasets, the extra analysis carried out showed that processing chains that undertake data smoothing produce comparable results despite the difference in the complexities of the algorithms used. An investigation into the key relationships between multiple datasets of both precipitation and vegetation was undertaken. This was an exploratory process that established the existence of a relationship between meteorological drought and agricultural drought in two aspects: High correlations coefficients (r) of between 0.67 and 0.76 and the existence of a lag of one month between vegetation and precipitation except for Mandera county that had less than a month of vegetation response to precipitation changes.

4.1.3 What is the relationship between the variables and drought severity?

The second research question seeks to find the relationship between each of the variables and drought severity. The end of this should be an objectively selected set of variables for the study. To answer this question, we proceed to document the results as follows: -

- Use of the study dataset to define the target variable based on the study methodology.
- The selection of one dataset and hence one set of variables from multiple competing datasets of the same variable.
- Investigate the relationship between the different target variables and different predictor variables

Definition of the target variable in the data follows the methodology section where two target variables are defined: drought severity based on VCI3M and drought impacts based on the percentage of children at risk of malnutrition as indicated by MUAC. Since the target was to predict the variables 1 month ahead, with the option to explore predictions 3-months ahead, the lag of the variables results in the number of records shown in Table 24 for the indicated periods.

Table 24: The record counts for each target variables (drought severity- VCI3M & drought effects- MUAC)

Target variable	Variable of Measure	Initial data period	Final data period	Total Records
Drought Severity	VCI3M	March 2001-Dec 2017	June 2001-Dec 2017	796
Drought Effects	MUAC	June 2008 – Dec 2017	July 2008 – Dec 2017	456

The two target variables were defined in terms of the VCI aggregated over 3 months period (VCI3M) and MUAC respectively as earlier indicated in section 3.6.1 on the methodical definition of the problem and as documented in Adede et al. (2019b).

The definition of the target variables indicates the availability of 796 and 456 training examples for the entire study. This number of records remains constant irrespective of the model definition as characterized by the number of model parameters. Since the most complex model has five (5) variables we used this case to establish the sufficiency of training data. We review data sufficiency in two contexts as follows: -

- One approach to determining the sufficient number of training examples was to use the rule of 10 for simple linear and the same 10 as the lower bound for non-linear prediction. Given the study had a maximum number of 5 variables, a minimum of 50 training records would be required. The 50 cases for training would indeed be unsuitable for the case of neural especially for the complexity of this study. We, however, used the 50 records as the base number of training records.
- The second approach was to define the neural network in terms of the weights that exist. In our case, we define this based on the most complex 5 variables case. We ensured that each connection and thus each weight to be learnt was defined as a model parameter. The number of parameters for the fully connected ANNs with the configuration 3-5-3-1and 5-3-3-1 deployed in this

study, therefore, gives a total of 52 and 34 edges respectively and with the bias node included in the count of the nodes as worked out from Equation 33.

$$E(n_1 - n_2 - n_3 \cdots n_i) = (n_{1*}n_2) + (n_{2*}n_3) + \cdots (n_i * n_{i-1}) \ldots\ldots \quad (33)$$

Extending the 10-rule to this problem still indicates the required size of training to be 520 for the prediction of drought severity and 340 for the prediction of drought effects.

The training dataset in this study, at 796 and 456 for the two target variables was thus deemed sufficient for the two approaches of analysing data sufficiency above. The number of training examples coupled with the tendency of ANNs to be robust even when training examples are few together with the other aspects of the methodology that includes cross-validation gives confidence in the sufficiency of the training examples.

4.1.3.1 Selection between TAMSAT and CHIRPS datasets

Variable Selection between multiple measures of the same concept follows from the methodology section where two target variables are defined as drought severity (VCI3M) and drought impacts (MUAC). For this study, the selection of variable was thus a choice between the TAMSAT and CHIRPS pairs of variables for the prediction of drought severity. For data reduction purposes, the most appropriate for the prediction of drought severity (VIC3M) between TAMSAT and CHIRPS was retained. The results for this investigation are documented as follows.

- ***Results of normality testing*** are presented for the non-SPI datasets. The SPI datasets are sampled off a normal distribution following on the approach in WMO (2012) and therefore conform to normality. As presented in Figure 4.5, all the other transformations of the CHIRPS and TAMSAT variables are shown not to conform to normality with all of them showing skewness. The test for normality was first done on non-normalized data as an initial test for conformity.

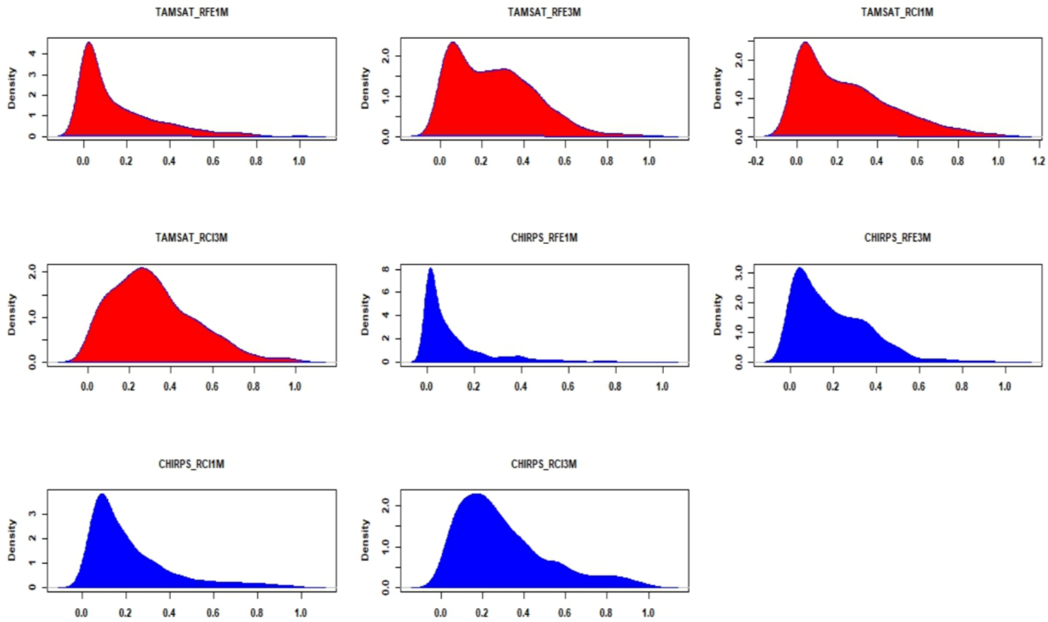


Figure 4.5: Density plots for the TAMSAT and CHIRPS non-normalized datasets

The visual plots in Figure 4.5 indicate the non-normality of all the non- standardized precipitation datasets. The non-conformance to normality was confirmed using the Shapiro Wilk test for significance as provided in Table 25.

Table 25: Shapiro-Wilk test on non-normalized CHIRPS and TAMSAT datasets

No	Variable	W	p-value
1	TAMSAT_RFE1M	0.8038	0.0000
2	CHIRPS_RFE1M	0.6956	0.0000
3	TAMSAT_RFE3M	0.9408	0.0000
4	CHIRPS_RFE3M	0.8987	0.0000
5	TAMSAT_RCI1M	0.9008	0.0000
6	CHIRPS_RCI1M	0.8191	0.0000
7	TAMSAT_RCI3M	0.9567	0.0000
8	CHIRPS_RCI3M	0.9065	0.0000

The hypotheses for the Shapiro-Wilk test for normality are as shown in Equation 34:

$$\begin{aligned} \boldsymbol{H}_0 &= \text{the population is normally distributed} \\ \boldsymbol{H}_1 &\neq \text{the population is not normally distributed} \end{aligned} \qquad \ldots\ldots\ldots (34)$$

From the above Shapiro-Wilk test, all the variables have p<0.05. The Null hypothesis was thus rejected and the alternative hypothesis that the population is not normally distributed was not rejected. Therefore, there was no evidence of normality for the variables.

Given that the predictive models were built using normalized variables, we additionally proceeded to do the test for normality on the normalized variables. The variables were normalized using the relative difference approach which scales the values in the [0,1] for [min, max] respectively. The density plots for the normalized variables is provided in Figure 4.6.

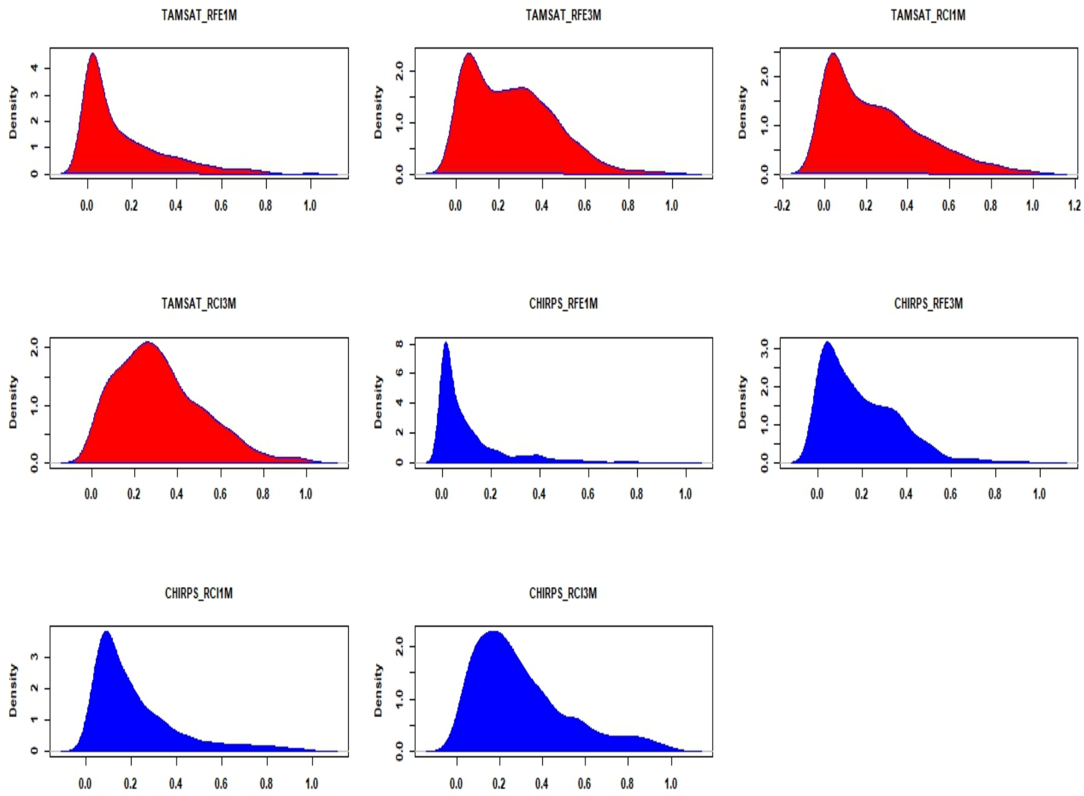


Figure 4.6: Density plots for the normalized TAMSAT and CHIRPS datasets.

The density plots indicate that the non-standardized precipitation datasets are non-normality distributed as supported by the corresponding Shapiro Wilk test in Table 26.

Table 26: Shapiro-Wilk test on normalized CHIRPS and TAMSAT datasets

No	Variable	W	p-value
1	TAMSAT_RFE1M	0.8038	0.0000
2	CHIRPS_RFE1M	0.6956	0.0000
3	TAMSAT_RFE3M	0.9408	0.0000
4	CHIRPS_RFE3M	0.8987	0.0000
5	TAMSAT_RCI1M	0.9008	0.0000
6	CHIRPS_RCI1M	0.8191	0.0000
7	TAMSAT_RCI3M	0.9567	0.0000
8	CHIRPS_RCI3M	0.9065	0.0000

The results of the analysis for normality clearly show that the normalized non-standardized precipitation datasets of both TAMSAT and CHIRPS are not normally distributed just as was the case for the non-normalized values.

The existence of the normally distributed SPI variables and the non-normally distributed variables of RFE & RCI implied the choice of methods for analysis had to follow on the non-parametric methods rather than the parametric methods. This is because the non-parametric approach to analysis does not rely on normality in the distribution of a variable's values. It is for this reason that, correlation analysis was done using Spearman's correlation rather than Pearson's correlation.

- ***Correlation analysis*** of the two datasets against the predicted variable of drought severity as defined by VCI3M was done using Spearman's rank correlation. The Spearman's rank correlation coefficients of the variables from both TAMSAT and CHIRPS are provided in Table 27. Generally, TAMSAT outperforms CHIRPS in four out of the six indicators. Overall, a mean of absolute correlation coefficients of 0.22 was posted by TAMSAT as compared to 0.18 for CHIRPS when analysed against the target variable for drought severity-VCI3M.

Table 27: Spearman's rank correlation for drought severity against TAMSAT and CHIRPS datasets

	RFE1M	**RFE3M**	**RCI1M**	**RCI3M**	**SPI1M**	**SPI3M**	**Mean**
TAMSAT	(0.02)	0.25	0.05	0.42	0.12	0.45	0.22
CHIPRS	(0.10)	0.13	0.07	0.32	0.09	0.36	0.18

Drought severity as indicated by VCI3M was shown to have a 1 month lagged relationship with the predictor variables. Table 28 shows the correlations of the predicted drought severity variable with the predictors from CHIRPS and TAMSAT lagged by 1 month.

Table 28: Spearman's rank correlation of VCI3M against 1-month lags of TAMSAT and CHIRPS

Variable	Type	Coeff
TAMSAT_RCI3M_lag1	TAMSAT	0.64
TAMSAT_SPI3M_lag1	TAMSAT	0.64
CHIRPS_RCI3M_lag1	CHIRPS	0.53
CHIRPS_SPI3M_lag1	CHIRPS	0.52
TAMSAT_RFE3M_lag1	TAMSAT	0.39
TAMSAT_SPI1M_lag1	TAMSAT	0.38
CHIRPS_RCI1M_lag1	CHIRPS	0.34
CHIRPS_SPI1M_lag1	CHIRPS	0.34
TAMSAT_RCI1M_lag1	TAMSAT	0.33
CHIRPS_RFE3M_lag1	CHIRPS	0.26
TAMSAT_RFE1M_lag1	TAMSAT	0.23
CHIRPS_RFE1M_lag1	CHIRPS	0.10

The superiority of 3-month indicators and that of TAMSAT variable are quite prominent with TAMSAT variables having an average correlation coefficient of 0.43 compared to 0.35 for TAMSAT variables.

- ***Step-wise forward regression*** uses a criterion on the F statistic to determine whether a variable can enter the model or not. The F statistic is influenced by the variables that are already in the model. This is akin to a greedy algorithm with the termination condition occurring when the next F to enter is not statistically significant at a specified threshold. This implies, like all greedy algorithms the possibility of use of a partial set of the variables. The result of step-wise regression indicates that two TAMSAT variables would not be selected despite the first two variables also being TAMSAT variables. The results of step-wise forward regression are as presented in Table 29.

Table 29: Step-wise forward regression of precipitation variables

Variable	Rank
TAMSAT_SPI3M	1
TAMSAT_SPI1M	2
CHIRPS_RFE1M	3
CHIRPS_RFE3M	4
CHIRPS_RCI3M	5
CHIRPS_RCI1M	6
TAMSAT_RFE1M	7
CHIRPS_SPI3M	8
TAMSAT_RCI3M	9
CHIRPS_SPI1M	10

- ***Bidirectional stepwise regression*** selected four variables: TAMSAT_SPI3M_lag1, CHIRPS_RCI3M_lag1, CHIRPS_SPI3M_lag1 and TAMSAT_RCI1M_lag1. Though TAMSAT_SPI3M ranked high, CHIRPS variables were also as competitive. TAMSAT ranked, consistently higher on the SPI that is widely documented in literature to be most appropriate for drought monitoring and is recommended by WMO(2012).
- ***The Akaike information criterion*** (AIC) was used in the study to estimates the quality of a model relative to others. The scope of models was defined to be ranging between the null model and the full model that is defined as the model of all the variables from precipitation. To ensure similarity in outputs, we exclude rows with missing values before the selection to ensure the use of a standard dataset. The results of the use of the AIC are as presented in Table 30.

Table 30: Akaike information criterion for variables selection

Variable	Df	Deviance	AIC
TAMSAT_SPI3M_lag1	1	19.596	-683.66
TAMSAT_RCI3M_lag1	1	20.532	-646.53
CHIRPS_RCI3M_lag1	1	24.155	-517.16
CHIRPS_SPI3M_lag1	1	26.279	-450.07
TAMSAT_RFE3M_lag1	1	27.075	-426.32
CHIRPS_RFE3M_lag1	1	29.731	-351.83
TAMSAT_SPI1M_lag1	1	30.078	-342.59
CHIRPS_RCI1M_lag1	1	30.763	-324.68
TAMSAT_RCI1M_lag1	1	30.944	-319.99
CHIRPS_SPI1M_lag1	1	31.456	-306.93
TAMSAT_RFE1M_lag1	1	32.341	-284.85
CHIRPS_RFE1M_lag1	1	33.161	-264.93

As shown in Table 30, TAMSAT again produced the first two variables ranked the highest although CHIRPS was competitive. It is key to note the performance of TASMAT in SPI as ahead on the selection of SPI from CHIRPS.

- ***Relative Importance of variables*** was used to determine the relative importance of variables fed into a linear model as a relative percentage as is used in Silber, Rosenbaum & Ross (1995). In the relative importance of variables approach, the R^2 is partitioned by averaging over the ordered list of performance which realized the results provided in Table 31.

Table 31: Relative importance of variables by Partitioned R^2

Variable	Relative Importance
TAMSAT_SPI3M_lag1	0.283
TAMSAT_RCI3M_lag1	0.197
CHIRPS_RCI3M_lag1	0.147
CHIRPS_SPI3M_lag1	0.096
TAMSAT_RFE3M_lag1	0.077
TAMSAT_SPI1M_lag1	0.050
CHIRPS_RFE3M_lag1	0.045
TAMSAT_RCI1M_lag1	0.029
CHIRPS_RCI1M_lag1	0.028
CHIRPS_SPI1M_lag1	0.024
TAMSAT_RFE1M_lag1	0.015
CHIRPS_RFE1M_lag1	0.009

The average relative importance also ranked SPI3M and RCI3M from TAMSAT higher than those from CHIRPS as evidenced in Table 31.

- ***The modelling approach to variable selection*** was used to also review the performance of the variables with a view to choosing the set considered to be more predictive. Support Vector Regression (SVR) and General Additive Model (GAM) models had the performances presented in Figure 4.7.

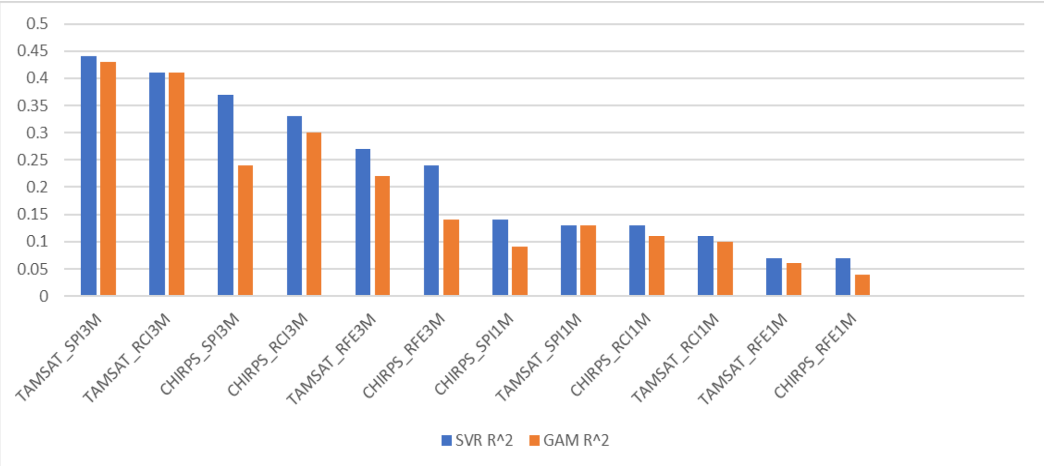


Figure 4.7: Performance (R^2) for both SVR and GAM models in variable selection.

The presented R^2 in Figure 4.7 is between drought severity (VCI3M) and the precipitation variables from either TAMSAT or CHIRPS. The single variable models were developed with the same model configurations using the in-sample dataset for the period 2001-2015.

From Figure 4.7, the SVR model was shown to generally outperform the GAM model for each of the variables except for TAMSAT_RCI3M and TAMSAT_SPI1M where similar performance is realised for the two models. The top performers for each of the GAM and SVR techniques are the two TAMSAT variables of SPI3M and RCI3M.

From the results of the analysis using the multiple methods, the study chose the TAMSAT dataset over the CHIRPS dataset. For the building of the models, only TAMSAT related precipitation datasets were used.

4.1.3.2 Relationship between non-precipitation variables and drought severity

Having presented the results of the investigation of the relationship between the precipitation variables and drought severity, we present, in this section, the relationship between the other variables with drought severity. This is presented using two approaches. First, using Spearman's rank correlation between the variables and drought severity as defined by VCI3M while the second is the use of modelling approaches.

A matrix of the coefficients of correlation between the pairs of all the non-precipitation datasets is provided in Figure 4.8.

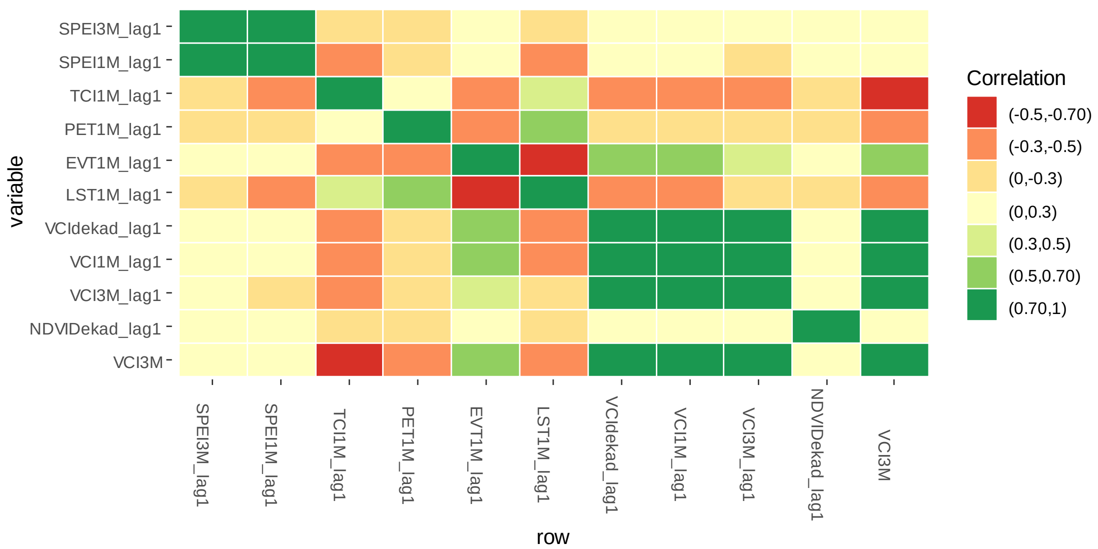


Figure 4.8: The correlation heat map for non-precipitation data for prediction of drought severity.

The correlational heat map shows a few highly correlated predictor variables as is expected especially amongst the vegetation variables. The 1 month lagged variables with the highest correlation to drought severity hence VCI3M includes the lags of the vegetation datasets like VCI3M_lag1, VCI1M_lag1 and VCIdekad_lag1. A summary of the correlation coefficients is as presented in Table 32.

Table 32: Correlation between non-precipitation data and drought severity (VCI3M)

Variable	Correlation with drought severity (VCI3M)
TCI1M_lag1	-0.58
LST1M_lag1	-0.45
PET1M_lag1	-0.34
NDVIDekad_lag1	0.16
SPEI1M_lag1	0.19
SPEI3M_lag1	0.28
EVT1M_lag1	0.59
VCI3M_lag1	0.82
VCI1M_lag1	0.88
VCIdekad_lag1	0.89

The temperature variables temperature condition index (TCI) and land surface temperature (LST) together with potential evapotranspiration (PET) are noted in Table 32 to have negative correlations with drought severity as was defined by VCI3M. On the other hand, the variables evapotranspiration (EVT) together with Standardized Precipitation Evapotranspiration Index (SPEI) and the vegetation variables have a positive correlation with drought severity. Given that drought severity is defined as less severe for higher VCI3M values, this inverse relationship between drought severity and the temperature and evapotranspiration variables is expected. There is the need to always interpret higher VCI3M vales as implying less drought.

It is the expectation that the best variables in the prediction of drought severity will see more vegetation and precipitation variables not leaving behind EVT and both TCI and LST.

4.1.4 Summary of investigation of relationships between variables and drought severity

The investigation of the relationship between the variables was a two-step process. First was the selection between TAMSAT and CHIRPS variables. The selection of variables was done for the two datasets rather than for all variables. This was because the approach of the study is a model space search rather than the selection of the best performer/ champion model.

The selection of variables as a modelling procedure was used for only the cases where multiple variables of the same kind of data existed. TAMSAT and CHIRPS are alternative data sources of the same indicators. Having both sources in a practical operational monitoring processing chain would be expensive in infrastructure and duplicative. We opine that choosing one of the datasets achieves a critical reduction in data volumes and also reduces efforts at data acquisition. We used multiple methods that posted mixed results in evidence for the choice between the TAMSAT and CHIRPS datasets. The results were generally geared towards the convergence on the superiority of TAMSAT in the prediction of drought severity as compared to CHIRPS. The modelling approaches of GAM and SVR and the correlational analysis of the variables in the prediction of drought severity are two approaches that lean towards the approaches used in model building. We, therefore, consider the evidence offered from GAM, SVR and correlational analysis as strong enough to lend credence to the selection of TAMSAT over CHIRPS from the pair of precipitation data sources.

We used this convergence of evidence to select TAMSAT as opposed to CHIRPS for inclusion in the model building process. Form the constellation of methods here-in, we conclude that even though both CHIRPS and TAMSAT are relatively shown to correlate with VCI3M at different levels, mostly with minor differences, most variables from TAMSAT tend to be ranked ahead of CHIRPS variables. We conclude that TAMSAT will be retained in the modelling process while CHIRPS excluded.

The investigation of the other variables and their correlation to drought severity values indicated quite strong correlations with drought severity. Table 33 provides a summary of the variables carried to the model building stage with their correlations with drought severity provided. All the precipitation variables provide are from the TAMSAT dataset.

Table 33: Summary of the correlation between the lagged predictor variables and future vegetation conditions (VCI3M).

Lagged Variable	Correlation with Drought Severity (VCI3M)	Type of relationship
TCI1M_lag1	−0.58	Moderate (-)
LST1M_lag1	−0.45	Low (-)
PET1M_lag1	−0.34	Low (-)
NDVIDekad_lag1	0.16	Weak (+)
SPEI1M_lag1	0.19	Weak (+)
RFE1M_lag1	0.23	Weak (+)
SPEI3M_lag1	0.28	Weak (+)
RCI1M_lag1	0.33	Low (+)
SPI1M_lag1	0.38	Low (+)
RFE3M_lag1	0.39	Low (+)
EVT1M_lag1	0.59	Moderate (+)
RCI3M_lag1	0.64	Moderate (+)
SPI3M_lag1	0.64	Moderate (+)
VCI3M_lag1	0.82	High (+)
VCI1M_lag1	0.88	High (+)
VCIdekad_lag1	0.89	High (+)

Note: The interpretation of the correlation follows the bands such that the magnitude shows the strength of the relationship while the sign shows the direction of the relationship. The absolute correlations are interpreted as follows:

- =0.0: No linear relationship
- 0.0 - <0.3: Weak
- 0.3 - <0.5: Low
- 0.5 - <0.7: Moderate
- ≥0.7: High to Very High

After the elimination of the 6 CHIRPS variables, we do not eliminate any of the other variables for three reasons:

- Combination of weak and strongly predictors has been documented to offer prediction models especially when their information value is such as to explain variations not explained by the strong predictors. The elimination of variables with weak correlations would then possibly eliminate variables independent from the variables highly correlated with the target variable.

- The combination of only high predictors in models leads to the problem of multi-collinearity since there is a chance the good predictors are linearly related to each other. We investigated the possibility of the occurrence of multi-collinearity in this study in section 3.7.7.2. The results of investigation of multi-collinearity were also documented in Adede et al. (2019a).
- The study was interested in investigating the performance of model ensembles in the prediction of both drought severity and drought effects. The first step to realizing model ensembles is the over-production of models to be followed by the selection of models. This investigation of ensembles provides for the opportunity to have different combinations of the variables evaluated in actual performance in models thereby justifying the non-elimination of variables exhibiting low correlation with the target variable.

4.2 Building and evaluating the performance of multiple models

To achieve the objective "*Build and evaluate the performance of multiple models for drought prediction using Artificial Neural Networks (ANN) and Support Vector Regression (SVR) as the case study Machine Learning methods"* we famulated the following research questions

- **RQ3:** What are the multiple models of both Artificial Neural Networks (ANN) and Support Vector Regression (SVR) that can be built for the prediction of both drought severity and drought effects? This question is handled in section 4.2.1
- **RQ4:** What is the performance of the ANN models as compared to SVR models in the prediction of drought severity? We handle this question in also handled in section 4.2.2
- **RQ5:** What is the performance of the ANN models as compared to SVR models in the prediction of drought effects? We handle this question in also handled in section 4.2.3

4.2.1 Building multiple Artificial Neural Network (ANN) and Support Vector Regression (SVR) models

To answer the question on "*What are the multiple models of both Artificial Neural Networks (ANN) and Support Vector Regression (SVR) that can be built for the prediction of both drought severity and drought effects?"* we did a model space search and identified all the models of ANN and SVR that could be built for the prediction of both drought severity and drought effects.

Prior to building the multiple models of both ANN and SVR and as outlined in the methodology, three key considerations were made: definition of the target variable in the data, normalization, sampling, model configuration and model space reduction. These key results from the pre-modelling steps are as presented here below:

(1) Definition of the target variable

The definition of the target variable saw the study structured into two predictive studies. First, for the prediction of drought severity and second for the prediction of drought effects. The target variables for drought severity and drought effects were chosen as VCI3M and MUAC respectively. In the methodology section, we defined MUAC as the percentage of children at risk of malnutrition. The Schema for the presentation of the results is provided in Figure 4.9.

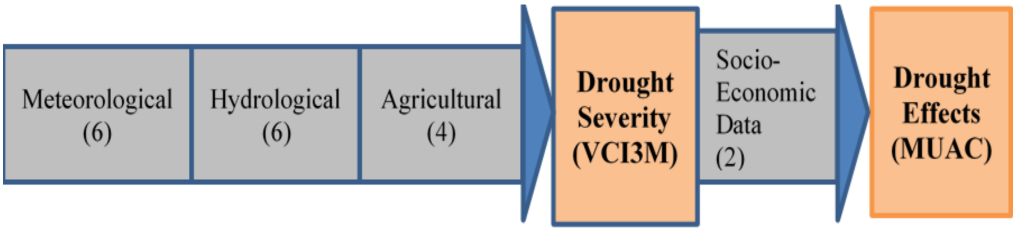


Figure 4.9: The model for the presentation of the results from the building of multiple ANN and SVR models.

Figure 4.9 indicates the flow of the modelling process as beginning with the prediction of drought severity and ending with the prediction of drought effects. The number of variables from each drought type is indicated. Before the lag of variables by up to 3 months in the study on drought severity, a total of 202 training examples per county had their target variable defined. With the lag, a total of 199 per county making for a total of 796 target variables were defined for drought severity. Similarly, the target variable for drought effects (MUAC) was set for 119 variables per county for a total of 476 records.

(2) Normalization

Normalization was done to ensure all variables are within a comparable range. The outputs of normalization were tested for correctness using correlation with the original dataset. The test for correctness established a correlation coefficient of one (r=1) between the normalized and non-normalized datasets. The confirmation of perfect correlation was done for all the variables and their normalized versions implying the transformation were correctly undertaken without loss of meaning to data. An example of the plot for the predicted variable and its normalized version is provided in Figure 4.10.

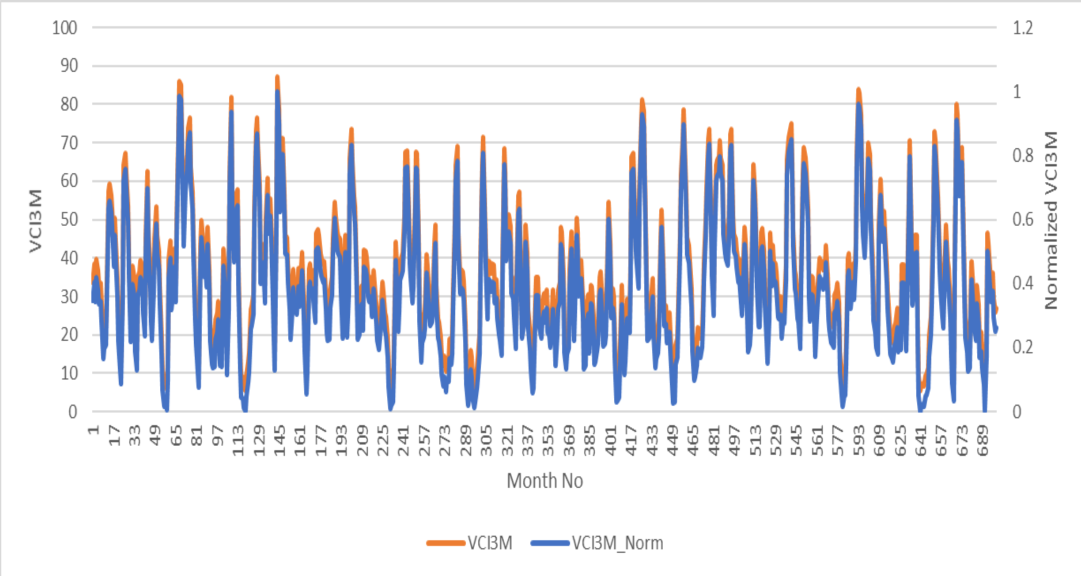


Figure 4.10: Results of test for correctness of normalization for the predicated variable. The correlation of the normalized variable against itself gives a correlation coefficient as well as an R^2 of 1.

(3) Sampling

The data split of the ANN process based on the methodology led to a three-way dataset split. The results of the data split are summarized in Figure 4.11 indicating an initial split of 90:10 on the in-sample and out-sample datasets and the subsequent split of the in-sample data into 70:30 for model training and model validation respectively.

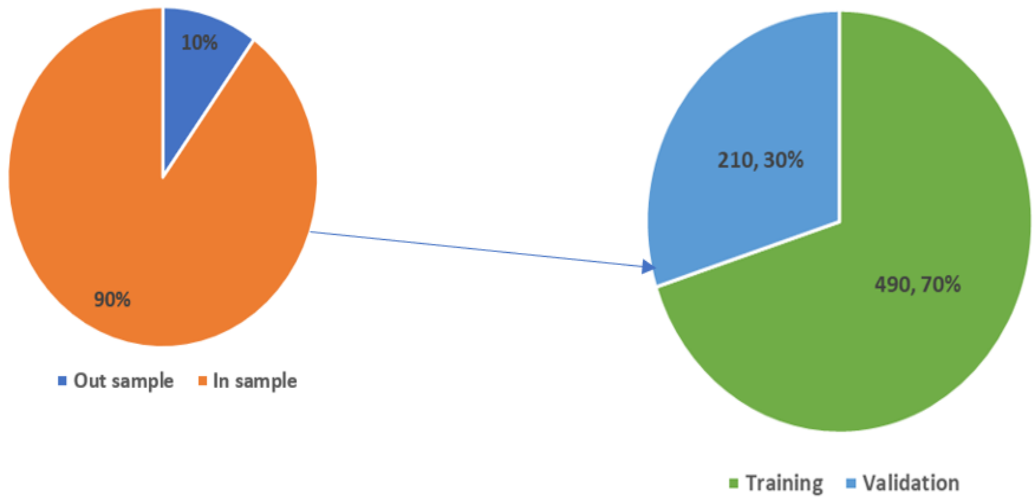


Figure 4.11: Results of the three-way division of sample data into in-sample and out-sample

(4) Model configuration

The investigation of the appropriate configuration for the ANN followed on the experimental approach provided for in the methodology. The results of some of the pivotal points of test done to achieve the configuration are presented in Table 34 Convergence was achieved with 10 rounds of experiments.

Table 34: Choice of model configuration

Configuration	**Converged**	**Proportion Models with $R^2 \geq 0.7$**
3-2-2-1	Yes	20%
3-3-2-1	Yes	23%
3-4-2-1	No	18%
3-4-3-1	No	46%
3-5-2-1	No	41%
3-5-3-1	Yes	53%

The results of experimentation using the rule of thumb mirror those of Huang (2003) that realises an optimal architecture with 5 nodes for the first hidden layer and 1 for the second. The prediction of drought effects retained the same configuration with the input layer upped to 5 for simplicity. This implied use of a less complicated architecture.

(5) Model Space Reduction

With the meteorological, hydrological and agricultural drought variables lagged 1 month to measure drought severity and subsequently to measure drought effects, it is noted the that the model space has an initial cardinality of ***65,535*** for drought severity models and ***262,143*** for drought effects models respectively as shown in Figures 4.12 and 4.13.

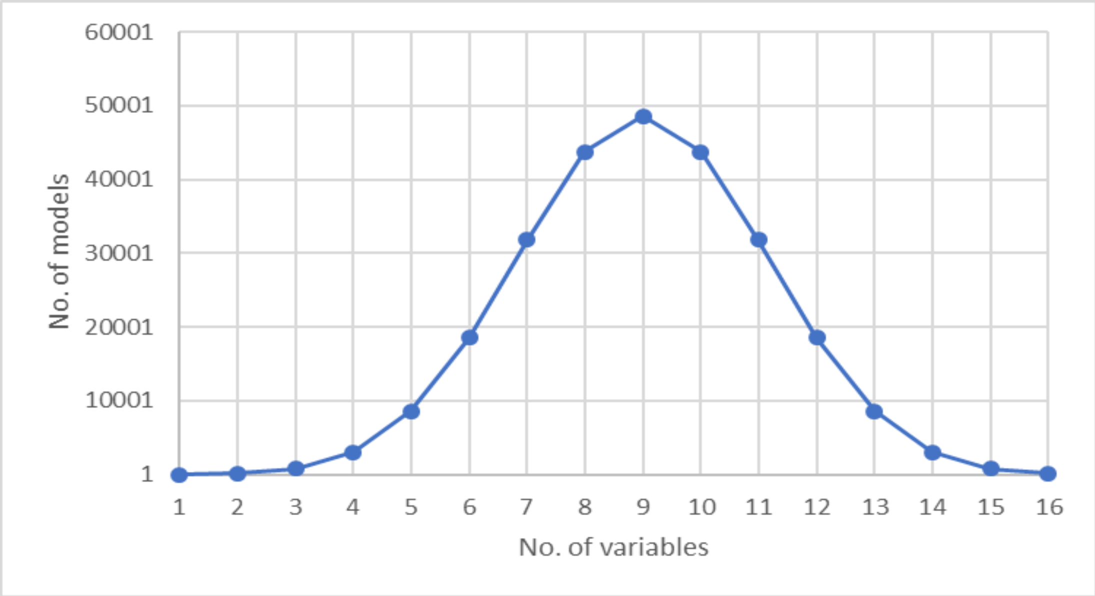


Figure 4.12: Model space of the drought severity prediction problem.
The figure indicates the number of models of each length for a total of 65,535 models.

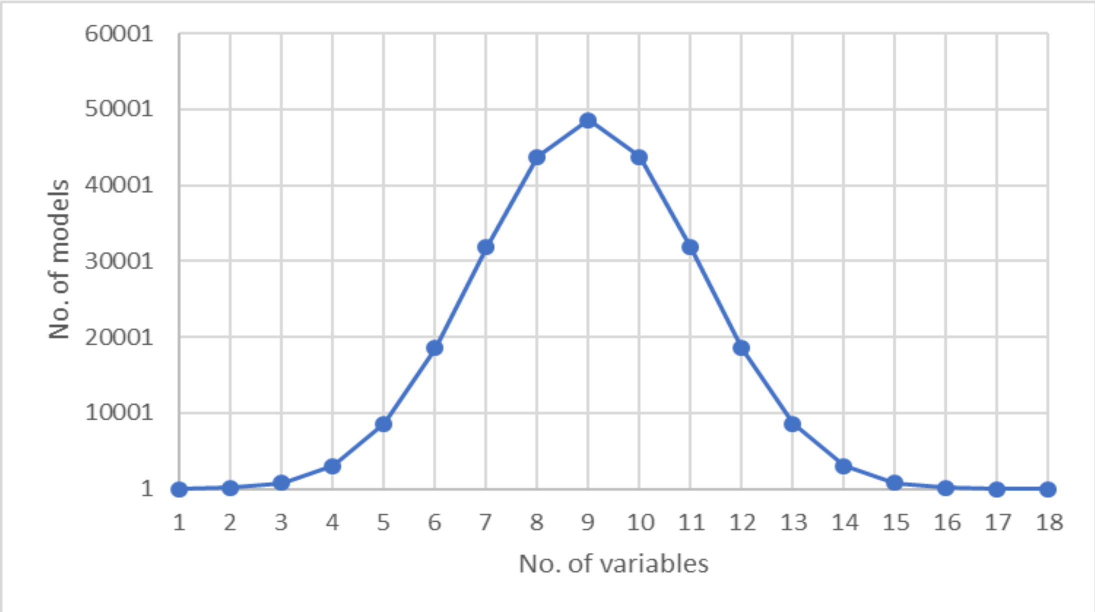


Figure 4.13: Model space of the drought effects prediction problem.
The figure indicates the number of models of each length from 1 to18 variables per model to realize a model space with a total of 262,143 models.

The cardinality of the model spaces for both drought severity prediction and drought effects prediction is quite enormous. Building multiple models from this space requires both experimentation and making of assumptions following the principle of no free lunch. The study made assumptions and established a cut off within which models were eliminated on the basis of their performance. Overall, the results of model space reduction are as shown in Figure 4.14.

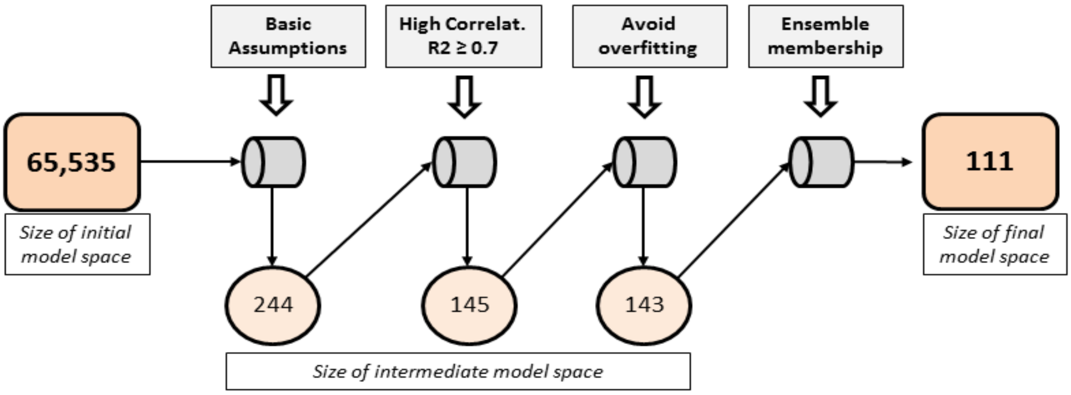


Figure 4.14: The model space reduction process

The assumption that adding multiple variables from the same grouping of variables only results into a marginal increase in model predictive power that was validated for soundness in Adede et al. (2019a) reduces the initial model space by over 99.6% to 244 models. It is these 244 models that were built using both the ANN and SVR techniques. Subsequently, for the prediction of drought effects the following combination of variables together with a recalibration of the models was done to achieve the building of multiple models for the prediction of future nutrition conditions as defined by MUAC:

- All the 244 drought severity models
- The 244 drought severity models' variables together with previous terms of trade (ToT)
- The 244 drought severity models' variables together previous MUAC values
- The 244 drought severity models' variables together with both terms of trade (ToT) and previous MUAC values

A total of 976 MUAC models were thus realized from the sub-step on building multiple models of the study methodology.

4.2.2 Performance of multiple ANN and SVR models in the prediction of drought severity

To answer the fourth research question (RQ4) on "*What is the performance of the ANN models as compared to SVR models in the prediction of drought severity?"* we analysed the performance of the 244 models built suing the ANN and SVR techniques.

This section presents the performance of both the ANN and SVR techniques in the perdition of drought severity 1 month ahead. The performance is presented as the performance in the training and validation datasets ordered by performance in the validation dataset. Performance is presented both for ANN and SVR techniques.

4.2.2.1 Performance of ANN in the prediction of drought severity in the training dataset (2001-2015)

From the total of 244 models that were subjected to the ANN modelling process for prediction of drought severity as indicated by the prediction of VCI3M 1 month ahead, 15 models representing 6.15% of the models were judged as overfitted and hence performed much less in model validation as compared to training by more than a 3% loss in performance. The distribution of the models based on their performance (R^2) is as indicated in Figure 4.15.

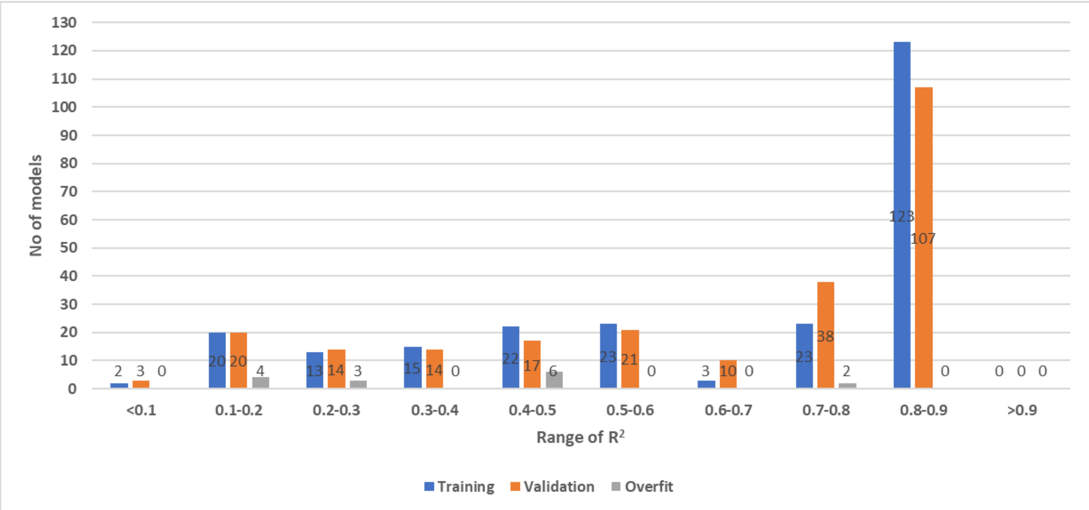


Figure 4.15: Performance (R^2) of the ANN models in the prediction of drought severity (VCI3M) 1 month ahead.

From Figure 4.15, it is remarkable that 145 models representing 59.43% of the models actually register an R^2 greater than 0.7 in the validation dataset and are thus considered to have acceptable predictive power in the prediction of drought severity. We provide an output of the performance of the first 30 ANN models in Table 35.

Table 35: Performance of the top 30 ANN models in training and validation ordered by descending performance in the validation dataset.

No	Model	R^2 (T)	R^2 (V)	Overfit Index	Overfit
1	VCI3M_lag1 + TAMSAT_RCI3M_lag1 + SPEI1M_lag1	0.81	0.86	0.05	0
2	VCIdekad_lag1 + TAMSAT_RFE1M_lag1 + TCI1M_lag1	0.87	0.86	-0.01	0
3	VCIdekad_lag1 + TAMSAT_RCI3M_lag1 + TCI1M_lag1	0.87	0.86	-0.01	0
4	VCIdekad_lag1 + TAMSAT_SPI3M_lag1 + TCI1M_lag1	0.87	0.86	-0.01	0
5	VCI1M_lag1 + TAMSAT_RCI3M_lag1 + TCI1M_lag1	0.87	0.86	-0.01	0
6	VCI3M_lag1 + TAMSAT_SPI3M_lag1 + SPEI1M_lag1	0.82	0.85	0.03	0
7	VCI1M_lag1 + TAMSAT_SPI3M_lag1 + TCI1M_lag1	0.87	0.85	-0.02	0
8	VCIdekad_lag1 + TAMSAT_SPI1M_lag1 + PET1M_lag1	0.87	0.85	-0.02	0
9	VCIdekad_lag1 + TAMSAT_SPI1M_lag1 + TCI1M_lag1	0.86	0.85	-0.01	0
10	VCI1M_lag1 + TAMSAT_SPI3M_lag1 + PET1M_lag1	0.87	0.85	-0.02	0
11	VCIdekad_lag1 + TAMSAT_SPI3M_lag1 + LST1M_lag1	0.86	0.85	-0.01	0
12	VCIdekad_lag1 + TAMSAT_SPI3M_lag1 + PET1M_lag1	0.87	0.85	-0.02	0
13	VCIdekad_lag1 + TAMSAT_RFE3M_lag1 + TCI1M_lag1	0.87	0.85	-0.02	0
14	VCI1M_lag1 + TAMSAT_RFE1M_lag1 + TCI1M_lag1	0.87	0.85	-0.02	0
15	VCI1M_lag1 + TAMSAT_SPI3M_lag1 + LST1M_lag1	0.86	0.85	-0.01	0
16	VCIdekad_lag1 + TAMSAT_RCI3M_lag1 + PET1M_lag1	0.86	0.85	-0.01	0
17	VCI1M_lag1 + TAMSAT_RFE3M_lag1 + TCI1M_lag1	0.86	0.85	-0.01	0
18	VCIdekad_lag1 + TAMSAT_RCI3M_lag1 + SPEI1M_lag1	0.86	0.85	-0.01	0
19	VCI3M_lag1 + TAMSAT_RCI3M_lag1 + SPEI3M_lag1	0.8	0.84	0.04	0
20	VCIdekad_lag1 + TAMSAT_SPI1M_lag1 + LST1M_lag1	0.86	0.84	-0.02	0
21	VCI1M_lag1 + TAMSAT_SPI1M_lag1 + PET1M_lag1	0.86	0.84	-0.02	0
22	VCI1M_lag1 + TAMSAT_SPI1M_lag1 + TCI1M_lag1	0.86	0.84	-0.02	0
23	VCIdekad_lag1 + TAMSAT_RCI3M_lag1 + LST1M_lag1	0.86	0.84	-0.02	0
24	VCI1M_lag1 + TAMSAT_RCI3M_lag1 + PET1M_lag1	0.86	0.84	-0.02	0
25	VCIdekad_lag1 + TAMSAT_RCI1M_lag1 + TCI1M_lag1	0.86	0.84	-0.02	0
26	VCIdekad_lag1 + TAMSAT_SPI3M_lag1 + SPEI1M_lag1	0.85	0.84	-0.01	0
27	VCIdekad_lag1 + TAMSAT_RFE1M_lag1 + LST1M_lag1	0.86	0.84	-0.02	0
28	VCI1M_lag1 + TAMSAT_RCI3M_lag1 + LST1M_lag1	0.86	0.84	-0.02	0
29	VCI1M_lag1 + TAMSAT_RFE1M_lag1 + LST1M_lag1	0.85	0.84	-0.01	0
30	VCIdekad_lag1 + TAMSAT_RFE1M_lag1 + SPEI1M_lag1	0.86	0.84	-0.02	0

R^2 (T)= Training R^2
R^2 (V)= Validation R^2

From the first 30 models (Table 35), a few observations emerge: -

- First, is the fact that there is no case of model over-fitting amongst the models given that no loss of more than 3% in performance between model training and model validation was recorded.
- Second, there exist 3 models amongst these top 30 models that we underfit and hence have a gain in performance. An extension of this analysis shows up to 4% of models as underfit amongst the top 100 models.
- Third, it is quite evident that the models have a little variance in their performance in the validation dataset. This is supported by Figure 4.16 that has a ranked plot of the models and shows surges in the performance in the training data as opposed to the validation dataset. This presentation is however expected since the ordering is by performance in training.

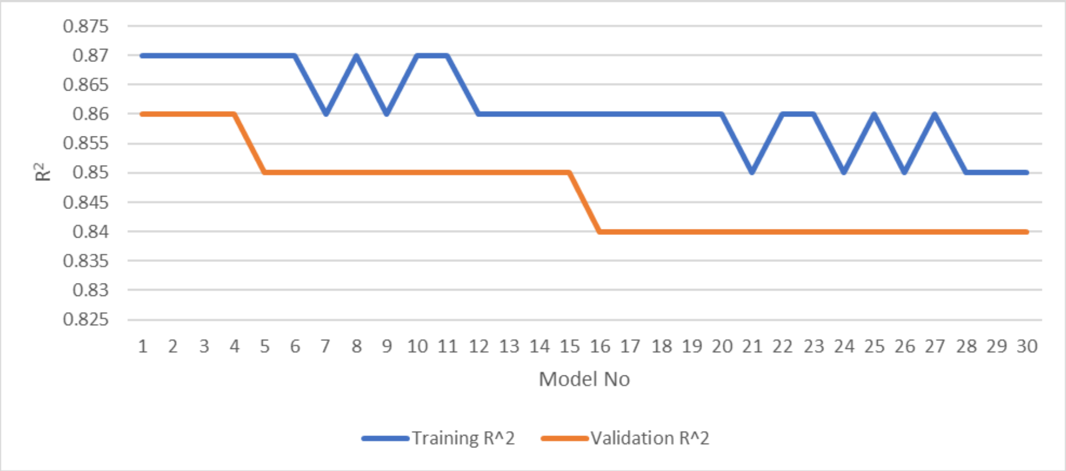


Figure 4.16: Ranked plot of performance of the first 30 models.

A plot of the measure of performance of each model based on Adjusted R^2 and RMSE for the 143 models chosen for model ensembling validates the expectation of the inverse relationship between the two measures of model performance. Despite the adjustment of R^2 to the Adjusted R^2, there still exists an inverse relationship between the two measures of performance on the data. This justifies the use of Adjusted R^2 as the measure of performance of choice. This correspondence is as shown in Figure 4.17.

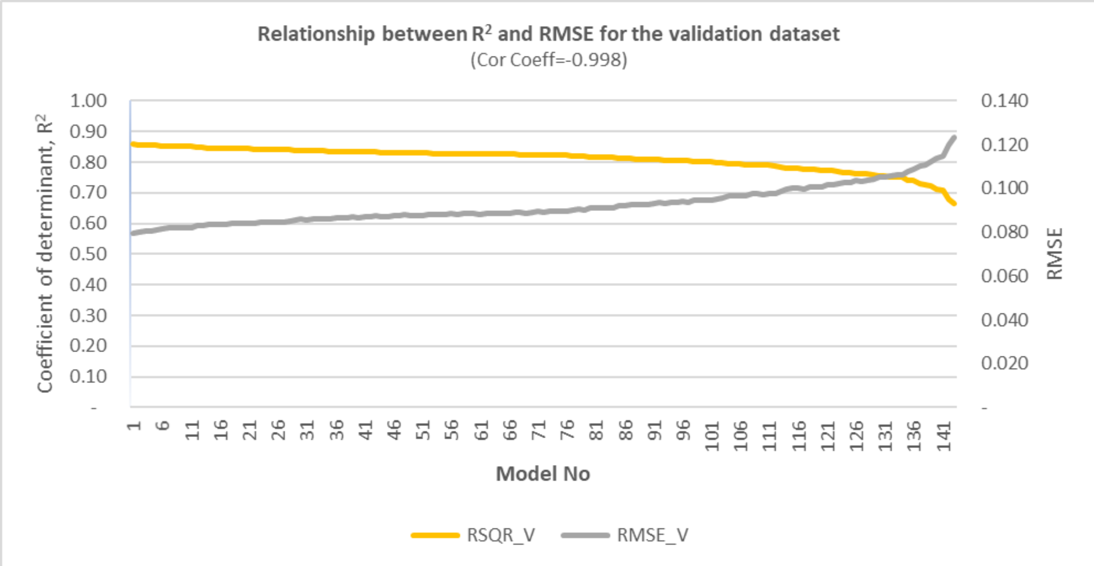


Figure 4.17: The relationship between Adjusted R^2 and RMSE on for the 143 ANN models.

There is shown a correlation coefficient of -0.998 between the Adjusted R^2 and RMSE in Figure 4.17 as expected. It, therefore, would make no difference if RMSE was used as the primary measure of model performance. The use of R^2 based on ease of interpretation is therefore expected to lead to the choice or rejection of similar models as would be the RMSE.

4.2.2.2 Performance of SVR in the prediction of drought severity in the training dataset (2001-2015)

The parameters used for SVR included an epsilon of 0.2, strategically chosen to avoid the problem of over-fitting. The cost parameter (C) was set to 32 as a result of an experimentation process that was based on the best averaged R^2 for the 244 models.

From the set of 244 SVR models, 145 of the 244 models representing 59.43% of the models had $R^2 \geq 0.7$ in the validation dataset. This performance is comparable to that of ANN models presented earlier. The analysis of model overfitting saw 21 models representing 8.61% of the 244 models judged to have been overfitted as shown in Figure 4.18.

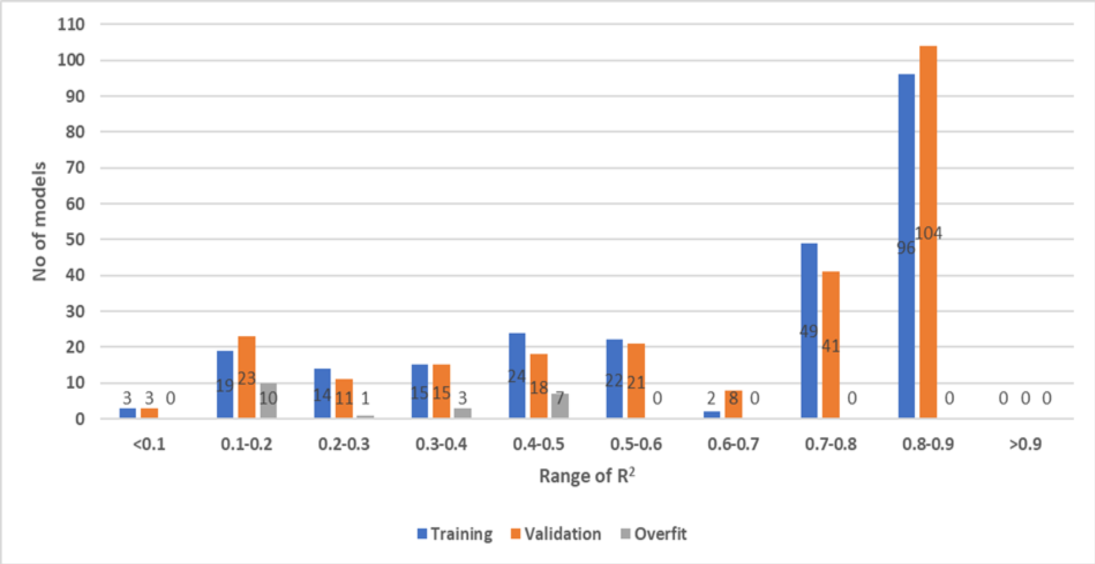


Figure 4.18: Performance of SVR drought severity models by R^2

The study presents, in Table 36, the performance in both training and validation of the top 30 SVR models ordered by performance (R^2) in the validation dataset.

Similar to the case ANN, the following facts emerge from Table 36:

- There is no occurrence of overfitting in the top 30 and top 100 models ordered by descending R^2 in the validation dataset.
- The fact that the occurrence of overfitting for the SVR technique is confined to the models with $R^2 \leq 0.5$.
- In a setting where the selection of models is based on the $R^2 \geq 0.7$ cut off, the problem of model over-overfitting would be confined to the ANN technique as compared to the SVR technique. This is so because there are two cases of model overfitting in the ANN process and none in the SVR process amongst all models with $R^2 \geq 0.7$. The tendency for models to suffer over-fitting in ANN with an increase in performance is for example documented in Mitchell (1997).

Table 36: Performance of the top 30 SVR models in training and validation ordered by descending R^2 in the validation dataset

No	Model	R^2 (T)	R^2 (V)	Overfit Index	Overfit
1	VCIdekad_lag1 + TAMSAT_SPI3M_lag1 + TCI1M_lag1	0.85	0.86	0.01	0
2	VCIdekad_lag1 + TAMSAT_RCI3M_lag1 + TCI1M_lag1	0.85	0.86	0.01	0
3	VCI3M_lag1 + TAMSAT_RCI3M_lag1 + SPEI1M_lag1	0.81	0.86	0.05	0
4	VCI3M_lag1 + TAMSAT_SPI3M_lag1 + SPEI1M_lag1	0.82	0.85	0.03	0
5	VCI1M_lag1 + TAMSAT_SPI3M_lag1 + TCI1M_lag1	0.85	0.85	0.00	0
6	VCIdekad_lag1 + TAMSAT_SPI3M_lag1 + PET1M_lag1	0.84	0.85	0.01	0
7	VCI1M_lag1 + TAMSAT_RCI3M_lag1 + TCI1M_lag1	0.85	0.85	0.00	0
8	VCIdekad_lag1 + TAMSAT_SPI1M_lag1 + TCI1M_lag1	0.85	0.85	0.00	0
9	VCIdekad_lag1 + TAMSAT_SPI3M_lag1 + LST1M_lag1	0.84	0.85	0.01	0
10	VCIdekad_lag1 + TAMSAT_RFE1M_lag1 + TCI1M_lag1	0.85	0.85	0.00	0
11	VCI1M_lag1 + TAMSAT_SPI3M_lag1 + PET1M_lag1	0.84	0.85	0.01	0
12	VCI1M_lag1 + TAMSAT_SPI3M_lag1 + LST1M_lag1	0.84	0.84	0.00	0
13	VCIdekad_lag1 + TAMSAT_SPI1M_lag1 + PET1M_lag1	0.84	0.84	0.00	0
14	VCI1M_lag1 + TAMSAT_RFE1M_lag1 + TCI1M_lag1	0.85	0.84	-0.01	0
15	VCI1M_lag1 + TAMSAT_SPI1M_lag1 + TCI1M_lag1	0.85	0.84	-0.01	0
16	VCIdekad_lag1 + TAMSAT_RCI1M_lag1 + TCI1M_lag1	0.84	0.84	0.00	0
17	VCIdekad_lag1 + TAMSAT_RCI3M_lag1 + LST1M_lag1	0.84	0.84	0.01	0
18	VCIdekad_lag1 + TAMSAT_RFE3M_lag1 + TCI1M_lag1	0.84	0.84	0.00	0
19	VCIdekad_lag1 + TAMSAT_RFE1M_lag1 + LST1M_lag1	0.84	0.84	0.00	0
20	VCIdekad_lag1 + TAMSAT_SPI1M_lag1 + LST1M_lag1	0.84	0.84	0.00	0
21	VCI3M_lag1 + TAMSAT_RCI3M_lag1 + TCI1M_lag1	0.83	0.84	0.01	0
22	VCIdekad_lag1 + TAMSAT_RCI3M_lag1 + SPEI1M_lag1	0.83	0.84	0.01	0
23	VCIdekad_lag1 + TAMSAT_SPI3M_lag1 + SPEI1M_lag1	0.84	0.84	0.00	0
24	VCIdekad_lag1 + TAMSAT_RCI3M_lag1 + PET1M_lag1	0.83	0.84	0.01	0
25	VCI1M_lag1 + TAMSAT_RCI3M_lag1 + LST1M_lag1	0.83	0.84	0.01	0
26	VCI1M_lag1 + TAMSAT_SPI1M_lag1 + PET1M_lag1	0.84	0.84	0.00	0
27	VCI3M_lag1 + TAMSAT_RCI3M_lag1 + SPEI3M_lag1	0.79	0.83	0.04	0
28	VCI1M_lag1 + TAMSAT_RFE3M_lag1 + TCI1M_lag1	0.84	0.83	-0.01	0
29	VCIdekad_lag1 + TAMSAT_SPI3M_lag1 + EVT1M_lag1	0.83	0.83	0.00	0
30	VCI1M_lag1 + TAMSAT_RCI1M_lag1 + TCI1M_lag1	0.84	0.83	0.00	0

R^2 (T)= Training R^2
R^2 (V)= Validation R^2

4.2.2.3 Comparative performance of ANN and SVR in the prediction of drought severity in the training and validation dataset (2001-2015)

Given that the study produced the same 244 models using both ANN and SVR techniques, it was imperative to compare the performance of the techniques. This is a valid comparison given that for the same model, the presentation of data in all the 10-folds used in both training and validation was the same across the techniques. The analysis of the performance of the pairings of the 244 ANN and SVR models is as presented in Figure 4.19.

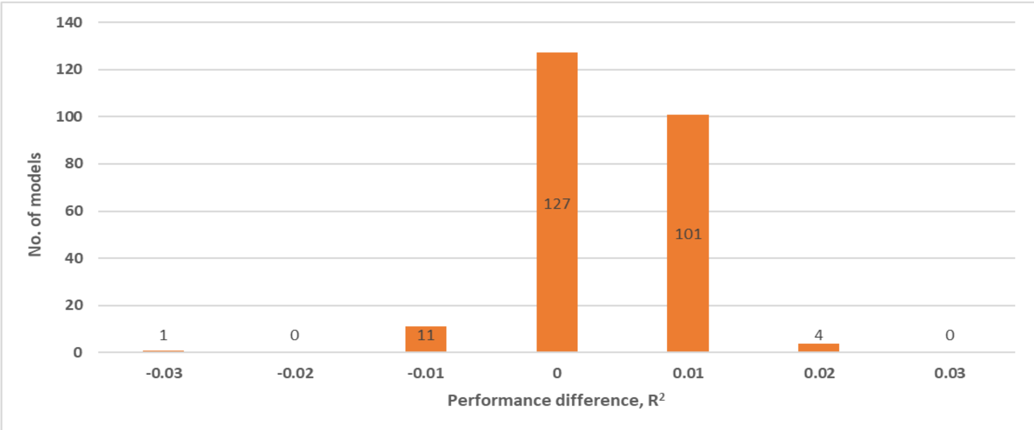


Figure 4.19: Comparative performance analysis between ANN and SVR model pairings.

The ANN and SVR techniques turned out competitive in model validation. 127 models representing 52% of the models posted similar performance. The ANN technique outperformed the SVR in 105 pairings representing 43% of the models while the SVR technique outperformed the ANN technique in 12 pairings making for 5% of the cases.

The analysis of the competitiveness of the ANN and SVR techniques was done using the summary statistics of minimum, maximum, average on the 143 models that had $R^2 \geq 0.7$ and were not overfit from the ANN process. As indicated in Table 37, the techniques were quite competitive in this set of 143 with similar scores across all the summary statistics.

Table 37: Summary of performance (R^2) for each technique in model validation.

Technique	Min	Max	Average	Range	StDev
SVR	0.71	0.86	0.81	0.15	0.03
ANN	0.71	0.86	0.81	0.15	0.03

The performance of the 143 models with $R^2 \geq 0.7$ is further elaborated in Figure 4.20. The techniques are further shown to be competitive in performance even on a percentage by percentage comparison.

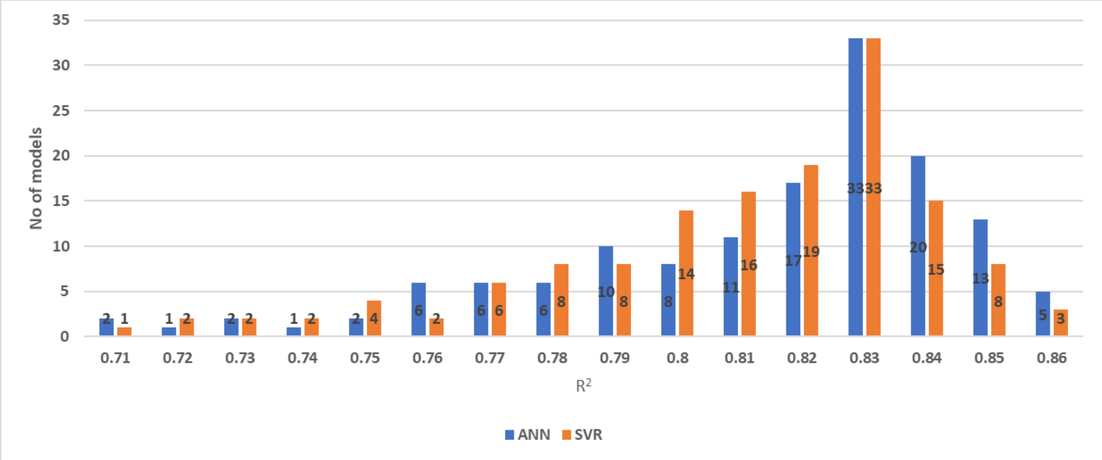


Figure 4.20: Comparative performance of ANN and SVR technique amongst models with $R^2 \geq 0.7$

Given no models with $R^2 \geq 0.7$ were overfitted from the SVR process, the choice of models for model ensembling was, therefore, a function of models from the ANN process. The selection of the appropriate models for ensembling was thus from the 143 ANN models paired with the corresponding SVR models of the same formula.

4.2.2.4 Comparative Performance of ANN & SVR models in the prediction of drought severity in the testing dataset (2016-2017)

The evaluation of the performance of the SVR and ANN models were also done using the test dataset that was designated as an out-sample dataset covering the period (2016-2017). We present the results of the performance of in the test dataset for all the 143 models that earlier had $R^2 \geq 0.7$ in model validation followed by a focused analysis on the performance of the champion models from both the ANN and SVR techniques.

In the overall performance of the models in the prediction of drought severity in the test dataset, the competitiveness of both techniques is clearly indicated by the minimum, maximum, average and range for both performance measures of R^2 and RMSE as shown in Table 38.

Table 38: Inter-technique performance analysis

Performance Measure	Statistic	ANN	SVR
R^2	Min	0.63	0.64
	Max	0.83	0.83
	Average	0.75	0.75
	Range	0.20	0.19
	StDev	0.03	0.03
RMSE	Min	0.08	0.08
	Max	0.11	0.11
	Average	0.09	0.09
	Range	0.04	0.03
	StDev	0.01	0.01

The SVR technique is judged competitive to the ANN technique and given the fact that it has fewer cases of over-fitting amongst the best performing models, it would be a technique of choice if the objective was to select the technique less prone to overfitting. From Table 38, an average ANN model will, therefore, perform comparably to a similar SVR model but with the tendency of the SVR model not to be overfitted. The distribution of the models in performance grouped by R^2 in Figure 4.21 indicates the performance of the ANN models as compared with SVR models in the testing dataset of the 143 models that had $R^2 \geq 0.7$ in the validation dataset. The ANN technique had 125 models with $R^2 \geq 0.7$ in the training dataset as compared to 132 models of SVR. It is therefore apparent that SVR suffers fewer cases of loss of performance and hence less overfitting as compared to ANN.

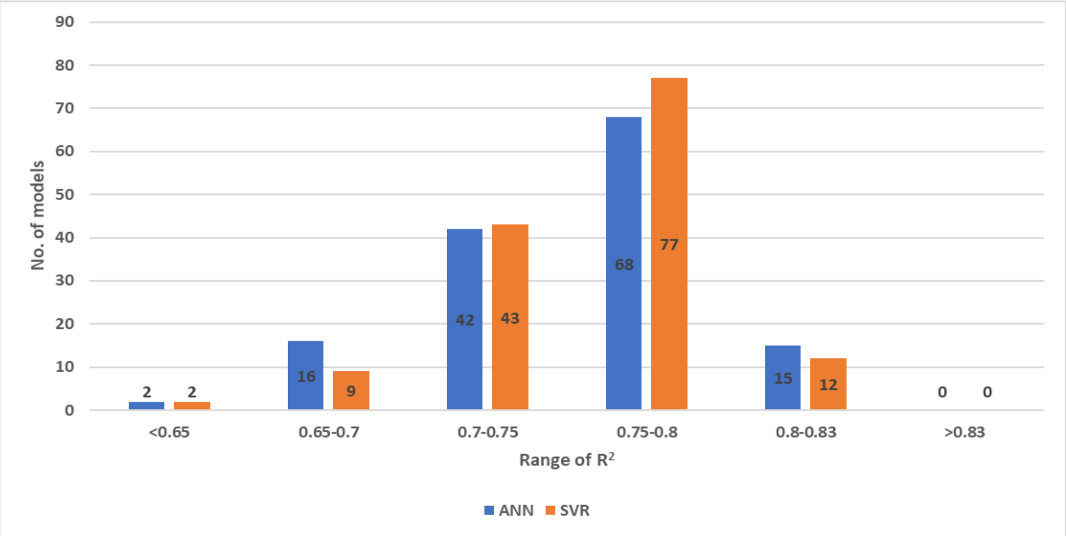


Figure 4.21: Drought severity performance of SVR versus ANN by grouped R^2

Given that most modelling approaches concentrate on the choice of a champion model for model scoring, we evaluated the performance of the best model from each of SVR and ANN techniques. The best model is chosen based on the performance in the prediction of VCI3M in the validation dataset. These two top models are summarised as presented in Table 39.

Table 39: Performance of the champion models for both ANN and SVR techniques

Technique	Champion Model	Validation R^2	Testing R^2
ANN	VCI3M + TAMSAT_RCI3M + SPEI1M	0.86	0.82
SVR	VCIdekad + TAMSAT_SPI3M + TCI1M	0.86	0.78

In the prediction of drought severity, the SVR champion and the ANN champion happen to be the different models. Though the models had the same performance in model training as indicated by R^2 in the validation dataset, their performance in the test dataset was at variance. While the ANN champion posted an R^2 of 0.82, the SVR champion posted an R^2 of 0.78 in the same test dataset. This is contrary to the earlier expectation as it shows more of the models from the SVR technique overfitting in the test dataset as compared to those from the ANN technique. In fact, it is interesting that for both the techniques there exist other models that out-perform the champion models with the best in each case posting an R^2 of 0.83 in the test data.

This loss of performance of champion models amounts to instability in performance and is the biggest limitation of the approaches that select champion models (Adede et al., 2019a). The performance of the ANN champion and the SVR champion in the test dataset in the prediction of future vegetation conditions (VCI3M) and hence drought severity is provided in Figure 4.22.

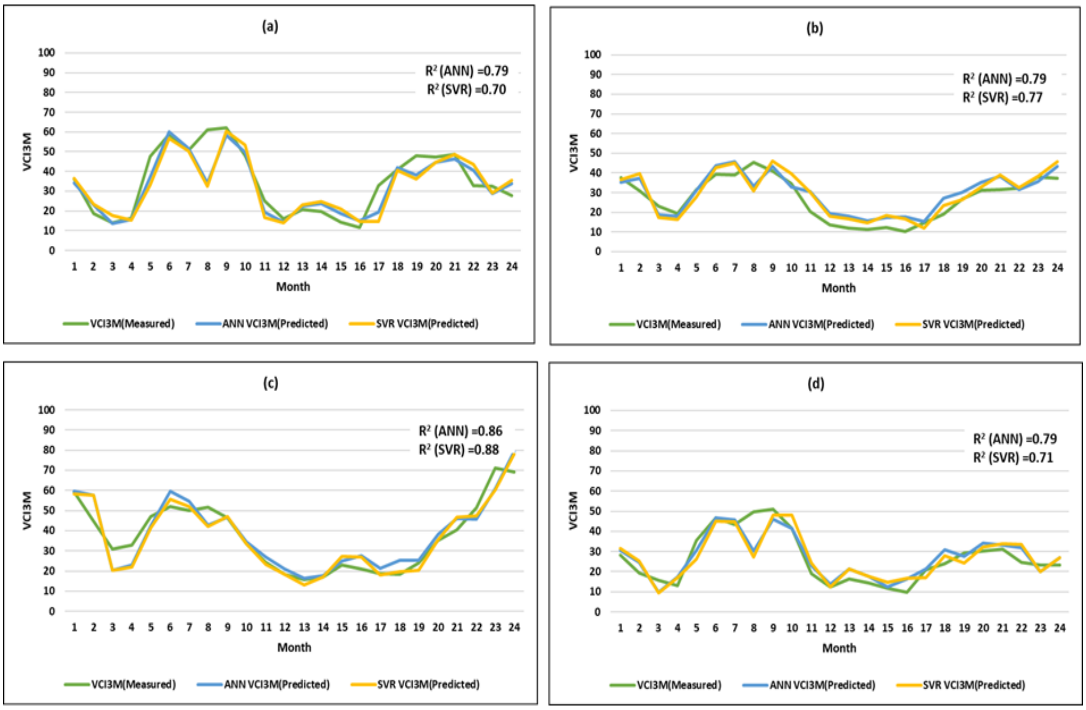


Figure 4.22: Comparative performance of the ANN and SVR champion models in the prediction of drought severity.

The plot in Figure 4.22 indicates the trend of the actual VCI3M values (green) and the predictions by the ANN champion (blue) and SVR champion (yellow) for the counties: (a) Mandera; (b) Marsabit; (c) Turkana and (d) Wajir. The best performance was posted for Turkana county by both the ANN champion and the SVR champion. Mandera and Wajir counties have a considerable wide gap in the performance between the ANN and the SVR champions of 9 and 8 percentage points respectively. The analysis at the county level, therefore, establishes that the utility of the ANN technique as better than that of the SVR technique.

The variance in the performance at the county level is possibly attributable to the champion models having been evaluated for performance on the overall dataset with the performance of R^2=0.82 judged satisfactory on the overall dataset. The performance at the county level is, however, acceptable considering that only one model would be learnt for the study area in this approach as compared to the alternative of having a model for each unit in the study area.

4.2.3 Performance of multiple ANN and SVR models in the prediction of drought effects.

To answer the fifth research question (RQ5) on "*What is the performance of the ANN models as compared to SVR models in the prediction of drought effects?*" we analysed the performance of the 976 models built suing the ANN and SVR techniques. In this section, we provide a comparative performance of the multiple ANN and SVR models in the prediction of drought effects as indicate by the proxy variable MUAC. The performance of the techniques in the prediction of drought effects is presented in two parts: Performance in model training and validation dataset (2008-2015) in section 4.2.3.1 and the performance in the test dataset (2016-2017) in section 4.2.3.2.

4.2.3.1 Comparative performance of ANN and SVR in the prediction of drought effects on the training dataset (2008-2015)

Using the same approach as that of drought severity, a total of 976 models were developed. The total number of models was as a result of the inclusion of ToT and MUA together in the models. In essence, there were: 244 models similar in variables formula to those of drought severity, 244 of drought severity and ToT, 244 of drought severity and MUAC and 244 of drought severity together with TOT and MUAC.

Compared to the prediction of drought severity, the prediction of drought effects records models that are judged to be poor performers even in the testing dataset both for ANN and SVR as shown in Figure 4.23.

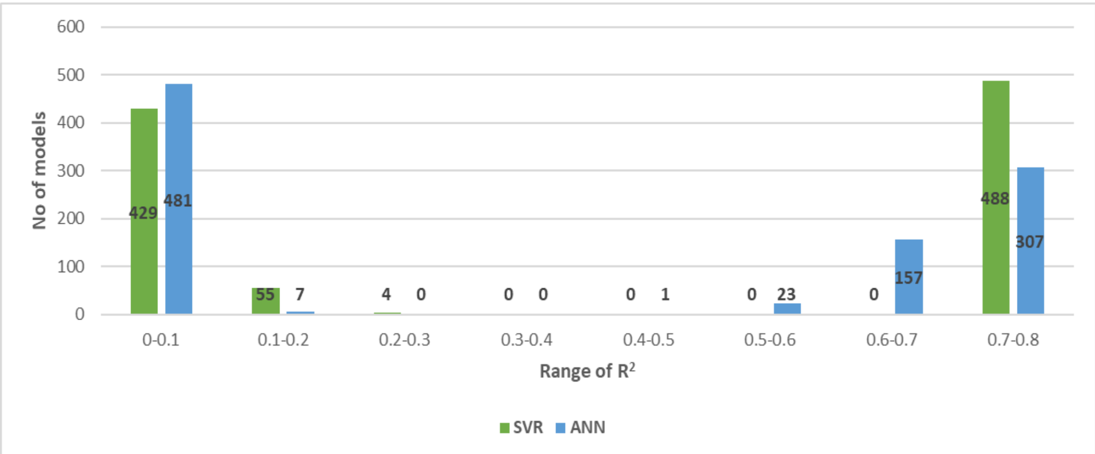


Figure 4.23: Performance of ANN versus SRV models in the prediction of drought effects

Despite both ANN and SVR approaches posting poor performance with 429 and 481 models having an R^2 less than 0.1, SVR out-performs ANN in the grouping of highly predictive models with R^2 between 0.7 and 0.8 by 488 models to 307 models.

From the set of models with $R^2 \geq 0.7$, the ANN technique reported 35 models as overfitting representing 27% of the models while none of the SVR models from this category was overfitting. This confirms the problem of overfitting as being pronounced for the ANN technique as compared to the SVR technique.

Furtherance of the comparative analysis of the performance of the ANN and SVR models in the prediction of drought effects at model training showed the SVR models as being competitive with the ANN technique with 488 to 464 models that performed better than chance and hence posted an $R^2>0.5$. This confirms the superiority of the SVR approach to the ANN approach in the prediction of drought effects as indicated by MUAC.

4.2.3.2 Comparative performance of the ANN and SVR techniques in the prediction of drought effects on the test dataset (2016-2017)

In the prediction of future nutrition conditions, the ANN and SVR models were run on the 24-month test dataset covering the period 2016-2017. The use of the data for the 2016-2017 period was the same as was done for the prediction of drought severity. A total of 96 data points was thus used to validate the performance of the ANN and SVR champions.

As is the case in the prediction of drought severity (VCI3M) 1 month ahead, the prediction drought effects (MUAC) 1 month ahead, also had two different champion models for both the ANN and SVR techniques. The SVR champion model had a total of 5 variables as opposed to the ANN champion model that had 3 variables. The ANN champion though outperformed by the SVR champion is a simple model since it had fewer variables and might thus be considered more appropriate in cases where model simplicity is a key factor. The overall performance across the entire dataset saw an R^2 of 7.4 for the ANN champion model as compared to 0.71 for the SVR champion model. This outcome is rather a contradiction, though not a surprise, in the sense that a method that was less competitive in producing more predictive models as indicated in Figure 4.19 actually produced the best overall model using the champion model approach. The utility of the SVR technique in generating multiple models over the ANN technique cannot be gainsaid.

The results of the performance analysis of the ANN champion and SVR champion at the county level in the prediction of drought effects (MUAC) 1 month ahead is as shown in Figure 4.24.

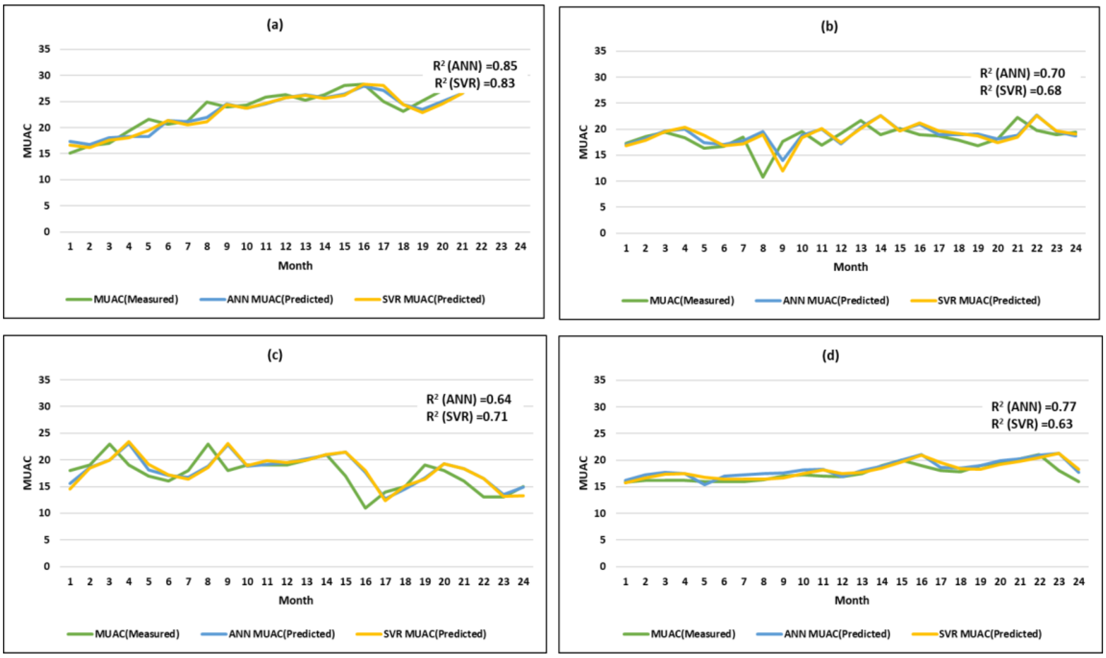


Figure 4.24: Performance of the ANN and SVR champion models in the prediction of drought effects.

Figure 4.24 confirms the closeness in the performance of the champion models across all the 24 months of test data for each of the counties: (a) Mandera, (b) Marsabit, (c)Turkana and (d) Wajir respectively. The SVR and ANN models do not show obvious specializations on any facets of prediction as they more or less retain the same trends across the peaks and troughs. The best performance is registered for Mandera county across both techniques with variances across the other counties. Notably, Turkana county sees the SVR champion model as outperforming the ANN champion in the prediction of drought effects.

In summary, the overall performance of the ANN and SVR models are quite close both in the prediction of drought severity and of drought effects for the entire test dataset. Variances are, however, noted when the analysis was carried at the county level. The choice between the champion models in both cases would result in the selection of the ANN technique in both instances since it produces the best overall model. Such a choice, however, loses out on the competitive performance of the SVR technique and especially its ability to produce many good performing models as compared to the ANN technique.

4.3 Building homogeneous and heterogeneous model ensembles

To achieve the objective "*Build and evaluate the performance of homogeneous and heterogeneous ensemble models of both ANN and SVR in the prediction of drought severity and drought effects"*, we famulated the following research questions

- **RQ6:** What is the performance of the Artificial Neural Networks (ANN) and Support Vector Regression (SVR) homogeneous ensemble models in the prediction of both drought severity and drought effects? This question is handled in section 4.3.2.
- **RQ7:** What is the performance of the ANN and SVR heterogeneous ensemble models in the prediction of drought severity and drought effects? This is handled in section 4.3.3.

The tendency of ANN to overfit as indicated in the results from RQ3 & RQ4 means it produced fewer models that are fit for purpose as compared to SVR technique, especially amongst the models that had $R^2 \geq 0.7$ in the validation dataset. The choice of the models for ensembling was therefore reduced to the selection of the appropriate ANNs for model ensembling and the subsequent pairing of the chosen ANN models with their SVR equivalents. Prior to the presentation of the results of the RQ6 & RQ7, we first handled the issues of model ensemble selection in section 4.3.1. The question of ensemble member selection remains critical in the effort to realize improved performance from model ensembling. We follow ensemble member selection with the review of the performance of the homogeneous ensembles in section 4.3.2 and the performance of the heterogeneous model ensembles in section 4.3.3.

4.3.1 Ensemble membership selection

To answer the research questions RQ6 and RQ7 above, for model ensembling we used only the models from the ANN technique that are indicated to have had $R^2 \geq 0.7$ and were not overfitted. The selection of models for ensembling, having been reduced to the question of selection of ANN models for ensembling was handled in two parts. First were the results presented in section 4.2.1(5) that saw the reduction of the model space from an initial 65,535 possible models to 244 models that were built using both the ANN and SVR techniques. This space was further reduced to 143 models that were judged to have an $R^2 \geq 0.7$ and were at the same time not overfitted for model ensembling.

Ensemble membership, also model pruning, has the results presented as the "selection" phase of the "over-produce" then select approach in model ensembling. The different reasons for model selection as documented in Mendes-Moreira et al. (2012) are to reduce computational costs, to increase prediction accuracy if possible and to avoid the problem of multi-collinearity.

From the 143 models that had an $R^2 \geq 0.7$ and were not overfitting, the construction of the ensemble membership faced two questions. First, was if all the 143 models were sufficient for the size of an ensemble and two was if there existed a smaller ensemble size that would perform the same, if not better than the 143. This question was answered following on the experimental process described in the methodology section and whose results are visualized in Figure 4.25.

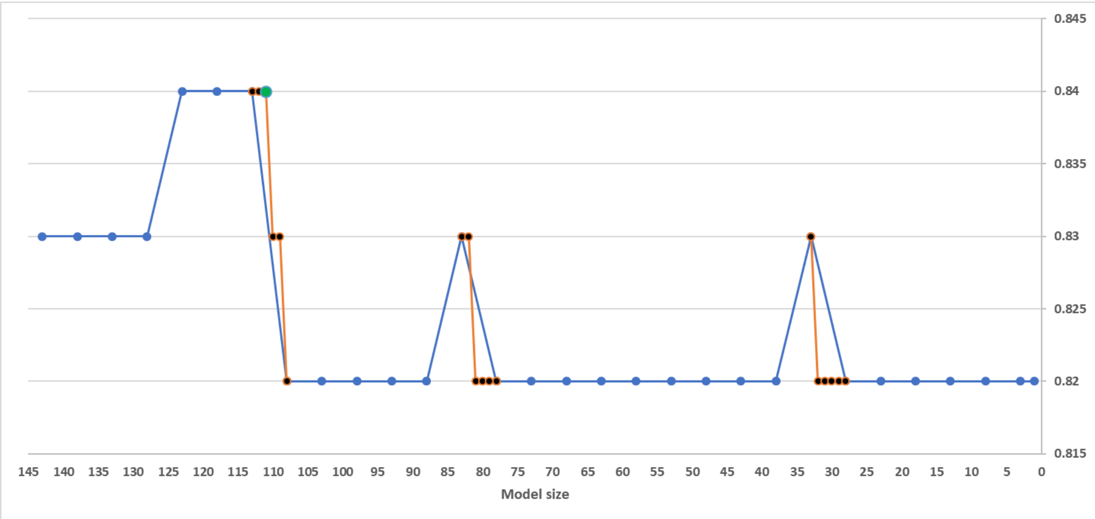


Figure 4.25: Ensemble membership selection showing the reduction from 143 models to 111 models.

Figure 4.25 shows the selection as eliminating the models in batches of 5 (blue line) while tracking any changes in ensemble performance. For example, the elimination of the first 3 batches, corresponding to the first 3 "dots" on the blue line from the left, sees no reduction in model performance. In fact, a further elimination, and hence reduction in ensemble size, realised an improvement in ensemble performance to an R^2 of 0.84 from 0.83 and stability of the ensemble performance thereafter till an overall ensemble size of 113 models. There was recorded a drop-in ensemble performance with the elimination of the batch of 5 from the ensemble size of 113 to 108. This marks the point at which forward substitution (orange lines) needed to be undertaken. This experimentation resulted in the best performing ensemble size of 111 models (highlighted as green dot) with an R^2 of 0.84. For convenience, the study proceeded with the iteration until all the batches were greedily eliminated and re-substitution was undertaken as shown in Figure 4.25.

The ensemble size of 111 was the best trade-off between ensemble size and ensemble performance. This is guaranteed to have a reduced computational complexity associated with the ensemble size choice whilst not losing ensemble performance. With the 111 models selected, we provide a plot of the weights in Figure 4.26 based on their performance in the validation datasets.

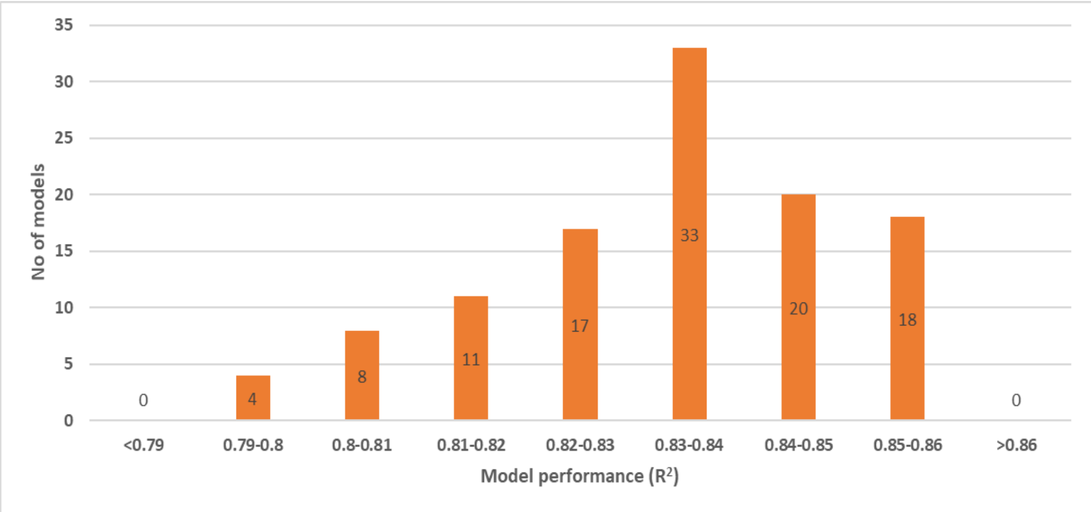


Figure 4.26: Distribution of performance of the selected ANN models for heterogeneous ensembling.

The 111 ANN models in Figure 4.26 chosen for model ensembling were paired with the same model definition of the SVR models for model ensembling. A total of 222 models were therefore used in building the model ensembles.

4.3.2 Homogeneous model ensembles

Homogeneous ensembles in the context of this study are ensembles in which all the models used in the ensemble are built using one technique. The ANN homogeneous ensemble was therefore exclusively made of models built using the ANN technique in the same way that homogeneous SVR ensemble was made entirely of models built using the SVR technique.

In this section, we present the results from ANN homogeneous ensembles and also from SVR homogeneous ensembles in the prediction of drought severity in sections 4.3.2.1 and 4.3.2.2 respectively. We also present the use of the ANN and SVR homogeneous ensembles in the prediction of drought effects in sections 4.3.2.3 and 4.3.2.4 respectively. We use the performance of the best champion model (R^2=0.83) as the basis of our comparison of performance. This section is a response to the sixth research question (RQ6) that poses the question on the performance of the homogeneous model ensembles in the prediction of both drought severity and drought effects.

4.3.2.1 Homogeneous ANN ensembles in the prediction of drought severity

The plot of the 111 ANN models selected for model ensembling for the prediction of drought severity had a ranked performance based on R^2 on the tests dataset as indicated in Figure 4.27. Any ensembling approach aims to realize an ensemble that outperforms the best base model with the highest R^2 of 0.82 as visualized in Figure 4.27. The best base model in this study is referred to as the champion model and specifically the ANN champion for the case of the ANN technique.

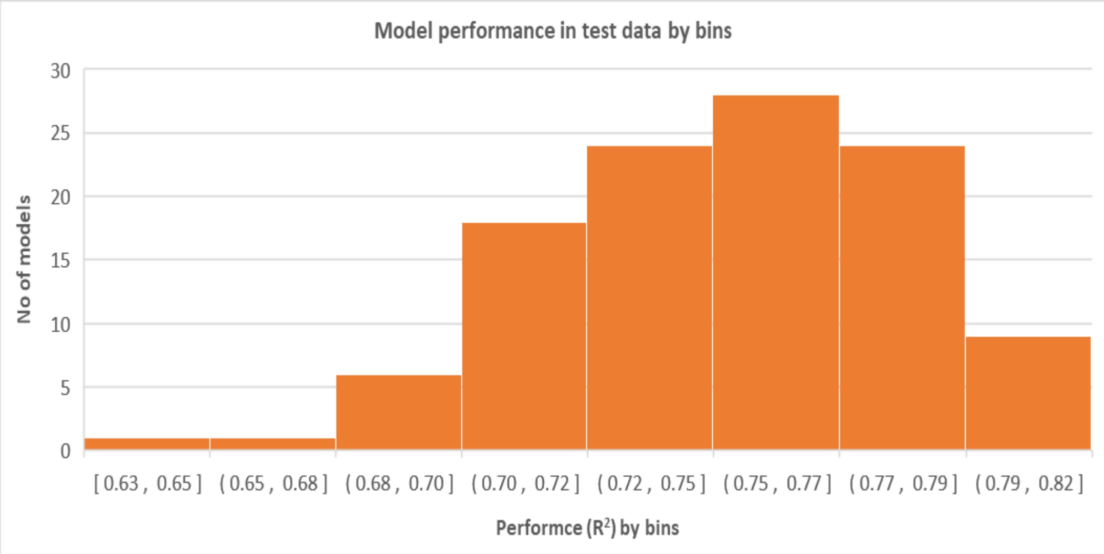


Figure 4.27: Performance of the ANN models in the test dataset

The non-ensembled models are shown to have lost performance in the testing dataset with the best performer and hence the champion model posting an R^2 of 0.82 from the earlier performance of 0.86 in the validation dataset. The performance of the base model for which ensemble performance was evaluated was, therefore, the champion model with an R^2 of 0.82 which is the ANN technique's best base model. Only about 6% of the models in Figure 4.28 are noted to either maintain their performance or to gain in performance in the test dataset.

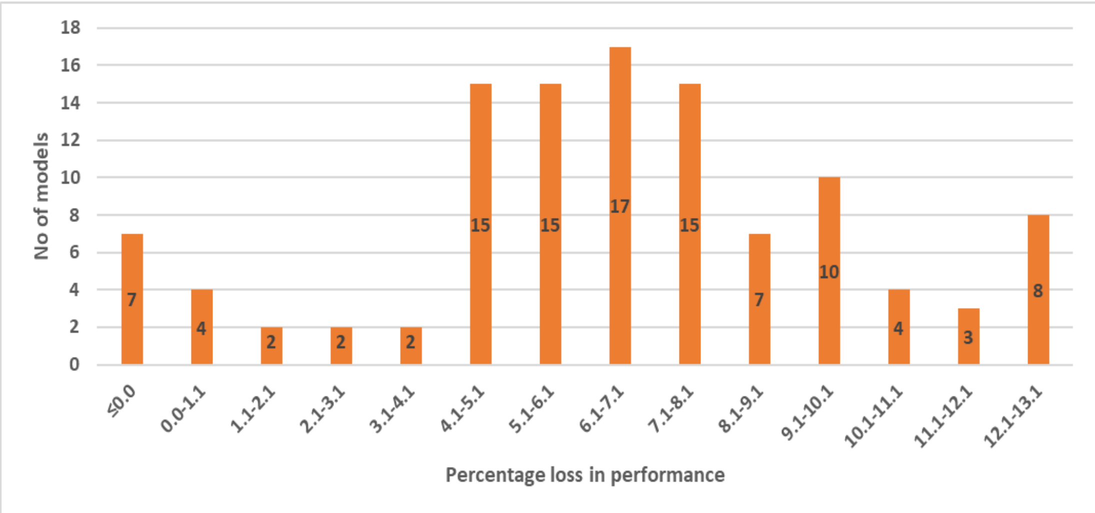


Figure 4.28: ANN models' loss of performance in the prediction of drought severity

We analysed the performance of the ANN homogeneous ensembles using two approaches:

- their performance in the prediction of future VCI3M values. We referred to this as their regression-based performance.
- their performance in the prediction of future VCI3M classes based on the classification in Table 42 earlier adopted from Klisch, Atzberger & Luminari (2015), Klisch & Atzberger (2016), Meroni et al. (2019) and used in Adede et al. (2019a) and Adede et al. (2019b). We referred to this as the performance of the ensembles in classification.

Regression based performance of homogeneous ANN ensembles in the prediction of drought severity

Table 40 provides a performance summary of the ANN ensemble models. The models are ensembles using the approaches outlined in the methodology: simple linear averaging, weighted average ensembling and ANN perceptron weighted ensembling.

Table 40: Summary performance of the ANN ensemble models in the prediction of drought severity (VCI3M) 1 month ahead

Model	MAE	MAPE	RMSE	R^2
ANN Champion	4.74	0.18	6.31	0.82
ANN Homogeneous Simple Average	4.43	0.17	5.96	0.84
ANN Homogeneous Weighted Average	4.35	0.17	5.85	0.85
ANN Homogeneous Stacked	3.40	0.13	4.40	0.91

From Table 40 it is evident that any form of ensembling is better that the use of the champion model that performs a little poorer as compared to the simple averaged, rank weighted and stacked model ensembles.

The performance gains offered by the simple weighting process and the rank weighting process on the champion model are 2 to 3 percentage point improvements respectively. Given the computational requirements to realise the 111 ensemble members, this might be a trade-off subject to the importance attached to the need for highly predictive models. For drought monitoring systems that inform resource allocation, the marginal improvement in performance is deemed as worth the investment in the accompanying computational complexity.

The approach in stacking used in this ANN approach learnt weights using an ANN perceptron. Given that all the ensemble members are ANN models, we characterise this as a homogeneous ensemble. The performance of the homogeneous ANN ensemble at an R^2 of 0.91 as shown in Table 40 is deemed to offer a significant improvement of 9 percentage points from the ANN champion that posted an R^2 of 0.82. The other metrics offer evidence of the superiority in the performance of the stacked model ensemble. The stacked ANN offers a clear-cut improvement in performance.

For strictly regression prediction, it is evidence enough that ANN perceptron learning of model weights in ensembles leads to the realization of more predictive models as compared to the use of the champion model or the averaging of the performance of the individual models. The superiority of the stacked ensembles over the averaging approaches is regardless of whether simple averaging or weighted averaging of the individual models in the validation dataset is undertaken.

An interesting aspect to the evaluation of the performance of the models is the performance of the homogeneous ensemble disaggregated at the county level as provided in Table 41.

Table 41: Performance (R^2) of the ANN homogeneous model ensembles disaggregated by county.

Approach	Mandera	Marsabit	Turkana	Wajir	Overall
ANN Champion	0.79	0.79	0.86	0.79	0.82
ANN Homogeneous Simple Average	0.78	0.86	0.88	0.80	0.84
ANN Homogeneous Weighted Average	0.79	0.86	0.88	0.81	0.85
ANN Homogeneous Stacked	0.93	0.87	0.89	0.93	0.91

The performance in Table 41 is presented for each of the model ensembling approaches of simple averaging, weighted averaging and model stacking. It is again apparent that the stacking approach guarantees the best improvement in performance as compared to other approaches. In-fact, there is a loss in performance in some instances like the case of simple averaging for Mandera county that recorded a one percentage point loss in performance as compared to the champion model approach.

The plot of the performance of the models in the prediction of drought severity in the test dataset is presented in Figure 4.29. A visual inspection of the plot of the actual values along the performance of the different ensembles indicates the closeness of the ANN stacked model to the actual values as compared to the other approaches to ensembling. The improvement in the performance of the ensemble approaches cannot be gainsaid save for the computation and process complexity that lead to their realization.

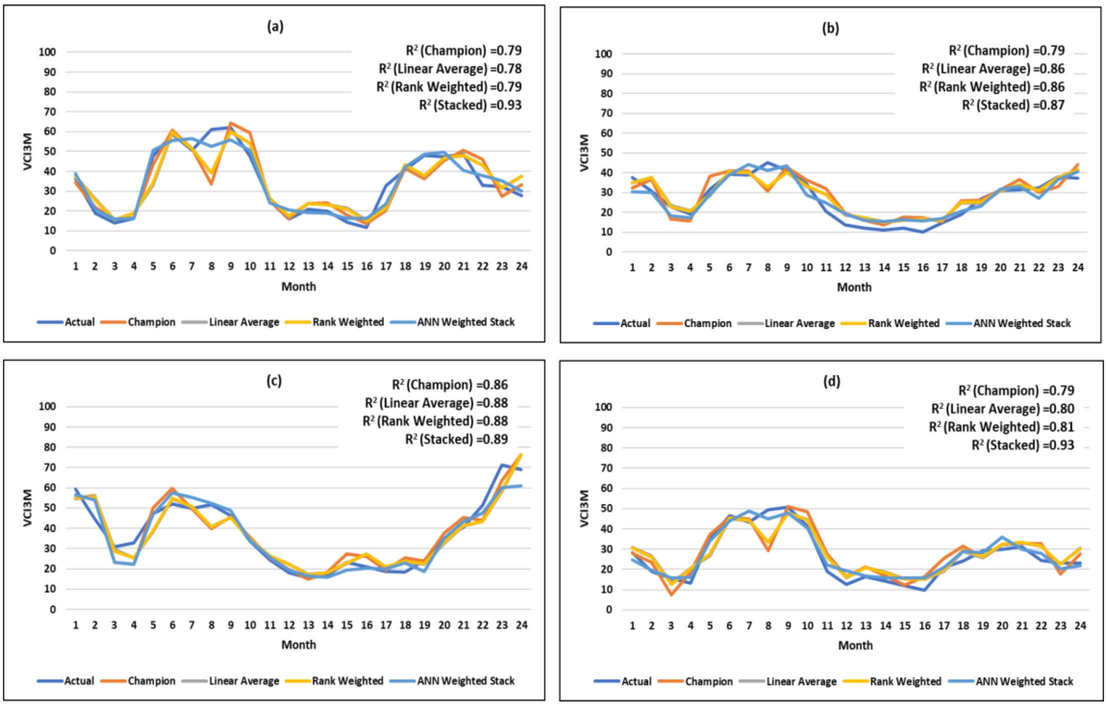


Figure 4.29: Plot of the actual values of VCI3M compared to the predictions from the champion model and the ensemble options.

The performance of the stacked ensemble in the test dataset (2016-2017) as shown in Figure 4.29 is not only close in trend to the actual values of VCI3M but is superior to the performance of the other homogeneous ensemble approaches. The performance at county level is: (a) Mandera (R^2=0.93); (b) Marsabit (R^2=0.87); (c) Turkana (R^2=0.89) and (d) Wajir (R^2=0.93).

Classification based performance of homogeneous ANN ensembles in the prediction of drought severity

For application in drought monitoring, most drought early warning systems (DEWS) convert the real value outputs to drought classes. The classes are easy to use in decision systems as they provide few decision points as compared to real value outputs from regression.

Towards this end of mirroring a decision system, the study compared the performance of the different homogeneous ANN model ensemble approaches but with the prediction formulated as class conversion problem from the outputs of regression. To define the classes on the VCI3M values used as the proxy for drought severity, we adopted the definition of vegetation deficit classes as shown in Table 42. The evaluation of performance in classification was based on five (5) classes defined on the VIC3M in the approach documented in Klisch, Atzberger & Luminari (2015), Klisch & Atzberger (2016), Meroni et al. (2019) and applied in Adede et al. (2019a). The vegetation deficit classes in this classification are: Above Normal, Normal, Moderate, Severe and Extreme vegetation deficit classes.

Table 42: Classification of drought based on vegetation deficit classes

VCI3M Limit Lower	VCI3M Limit Upper	Description of Class	Drought class
≤0	<10	Extreme vegetation deficit	1
10	<20	Severe vegetation deficit	2
20	<35	Moderate vegetation deficit	3
35	<50	Normal vegetation conditions	4
50	≥100	Above normal vegetation conditions	5

Note: The classification table was earlier presented in Table 16. It is repeated here for convenience in interpretation.

Using the thresholds in Table 42 and subjecting the monthly data to the classification of vegetation conditions, we obtained the drought classes that were denoted to range from 1-5 for the Extreme, Severe, Moderate, Normal and Above normal conditions respectively.

The confusion matrix that results from the conversion of the outputs of regression to drought classes is as shown in Table 43 for the ANN champion and each of the ANN homogeneous ensemble approaches.

Table 43: Confusion Matrix for the champion model and each of the homogeneous model ensemble approaches

	(a) ANN Champion						(b) Simple Average						(c) Weighted Average						(d) Stacked				
	Actual						Actual						Actual						Actual				
Predicted		1	2	3	4	5		1	2	3	4	5		1	2	3	4	5		1	2	3	4	5
	1	0	1	0	0	0	1	0	1	0	0	0	1	0	1	0	0	0	1	0	1	0	0	0
	2	1	18	9	0	0	2	0	16	12	0	0	2	0	16	12	0	0	2	0	22	6	0	0
	3	0	3	24	4	0	3	0	1	27	3	0	3	0	1	27	3	0	3	0	5	24	2	0
	4	0	0	4	19	2	4	0	0	5	17	3	4	0	0	6	16	3	4	0	0	2	20	3
	5	0	0	1	3	7	5	0	0	0	4	7	5	0	0	0	4	7	5	0	0	0	2	9

Table 43 presents the contingency tables for the champion model approach (a) and the homogeneous model ensemble approaches of (b), (c) and (d) for the simple average, weighted average and stacked model ensembles respectively.

From the individual contingency tables, the measures of Accuracy and Kappa, Sensitivity and Specificity were calculated using the approach in Kuhn (2008) for the multi-class classification problem. The results of the above measures of classification performance are as provided in Table 44– 46 for the measures in the listed order.

Table 44: Performance by overall accuracy and Kappa for each model ensemble approach

Model	**Accuracy (%)**	**Kappa**
ANN Champion	71	0.60
ANN Homogeneous Simple Average	70	0.58
ANN Homogeneous Weighted Average	69	0.57
ANN Homogeneous Stacked	78	0.70

Table 45: Performance by Sensitivity for each model ensemble approach

Model	Class				
	1	2	3	4	5
ANN Champion	0	0.82	0.63	0.73	0.78
ANN Homogeneous Simple Average	NA	0.89	0.61	0.71	0.70
ANN Homogeneous Weighted Average	NA	0.89	0.60	0.70	0.70
ANN Stacked	NA	0.79	0.75	0.83	0.75

Table 46: Performance by Specificity for each model ensemble approach

Model	Class				
	1	2	3	4	5
ANN Champion	0.99	0.86	0.88	0.91	0.95
ANN Homogeneous Simple Average	0.99	0.85	0.92	0.89	0.95
ANN Homogeneous Weighted Average	0.99	0.85	0.92	0.88	0.95
ANN Stacked	0.99	0.91	0.89	0.93	0.98

From the presentation of performance in Tables 44-46, it is clear that posed as a classification problem, the stacked model ensemble out-performs the other approaches in model accuracy. Given that Kappa takes into account the possibility of the agreement occurring by chance, the weighted approach to model ensembling has the lower probability of agreement by chance as compared to the stacked model ensemble that is higher in classification accuracy. The Kappa measure, in this case, poses the "*paradox of high agreement and lower Kappa". This paradox informs the use of the True Positive Rate (TPR) that* is also referred to as the *Sensitivity/Recall/ Probability of detection* and *Specificity (True negative rate)* with the view to optimization as presented in the multi-class receiver operating characteristics curve (ROC).

It is apparent from the results in Table 44 that the model stacking approach to ensembling offers the best accuracy in the prediction of drought severity at 78% as compared to the base performance of 71% of the ANN champion model. Notedly, the simple average and weighted average approaches to homogeneous ANN ensembles record marginal losses in performance. A county by county disaggregated analysis of the performance of the approaches is presented in Table 47 showing the highest range in performance by the stacked ensemble approach as compared to the champion and the averaging approaches. The stacked ANN ensemble clearly outperforms the other model ensembles in the prediction of drought severity.

Table 47: Classification accuracy of the ANN homogeneous ensembles for each county

Approach	**Mandera**	**Marsabit**	**Turkana**	**Wajir**	**Overall**
ANN Champion	0.71	0.75	0.71	0.67	0.71
ANN Homogeneous Simple Average	0.67	0.83	0.67	0.63	0.70
ANN Homogeneous Weighted Average	0.67	0.79	0.67	0.63	0.69
ANN Homogeneous Stacked	0.79	0.88	0.75	0.71	0.78

Given that both sensitivity and specificity as presented in Table 45 and Table 46 respectively are class-dependent, better visualization of the results would be through the use of the ROC curve as presented in Figure 4.30 following the approach in Hand & Till (2001).

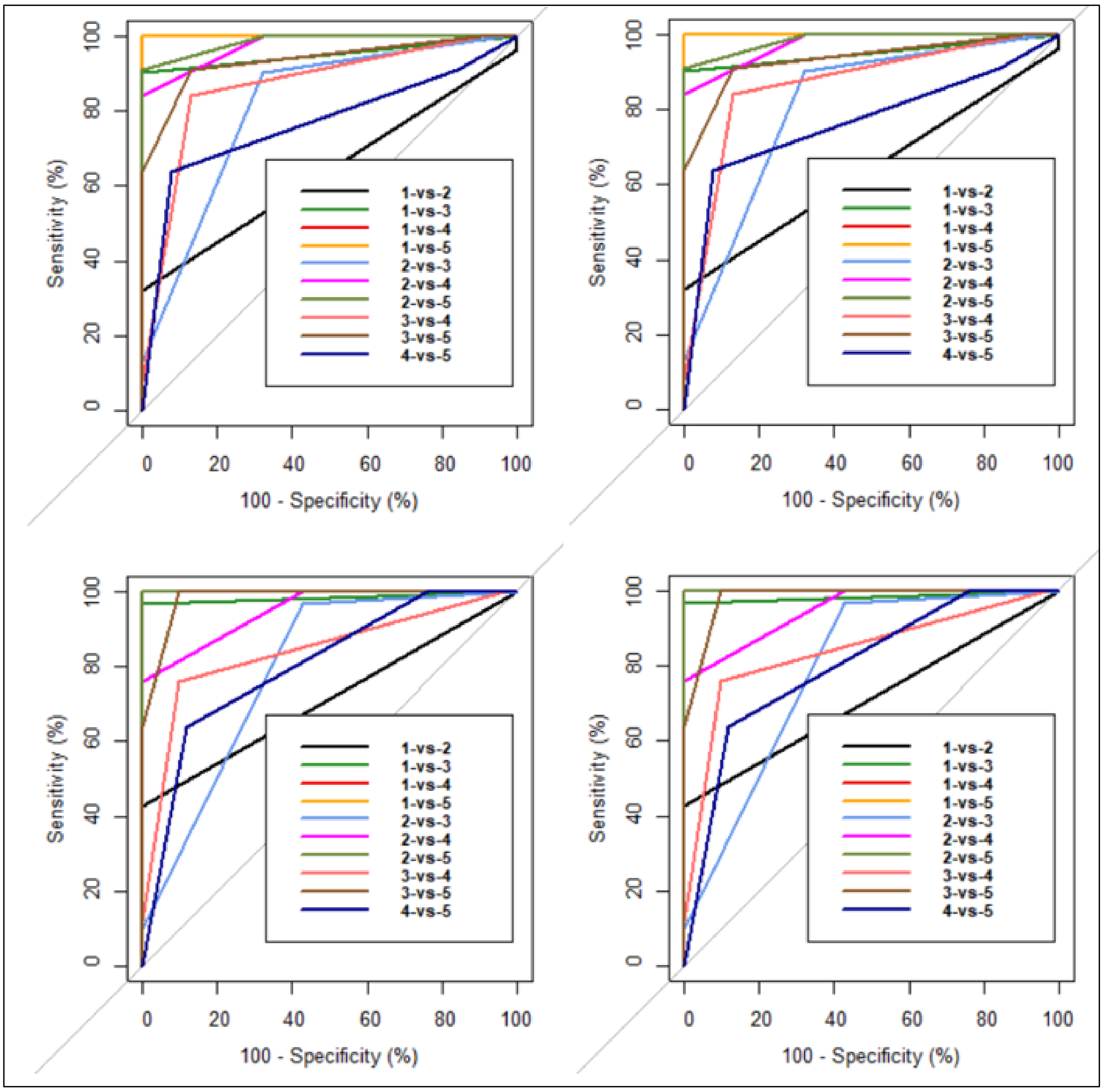


Figure 4.30: The multi-class ROC plot for the ANN champion and the homogeneous model ensembles.

The performance was based on the area under the ROC curve (AUROC) and was calculated using the one-versus-all approach. The ROC plot indicates the ranking of the models as classifiers with a trade-off between sensitivity and specificity. The performance posted, clockwise from upper left recorded: an AUROC of 89.37 for the

ANN champion, 90.83 for the simple average ensemble, 90.63 for the weighted average ensemble and the best performance of 91.26 for the stacked homogeneous ensemble for the ANN technique. The definitive advantage offered by the homogeneous ANN stacked ensemble is evident with an area under the curve (AUC) of 91.26% as compared to that of the champion model at 89.37%.

4.3.2.2 Homogeneous SVR ensembles in the prediction of drought severity

The performance of the homogeneous ensembles built from the SVR technique is discussed following on the same order as that used for the homogeneous ANN ensembles in both regression and classification.

Regression based performance of homogeneous SVR ensembles in the prediction of drought severity

The SVR models, just like the ANN models, lost performance in the test dataset except for 4 models that marginally registered gains in performance. The loss of performance amongst the SVR models in the prediction of drought severity is as depicted in Figure 4.31. This loss of performance is based on the difference between R^2 in the validation dataset and that of the test data set.

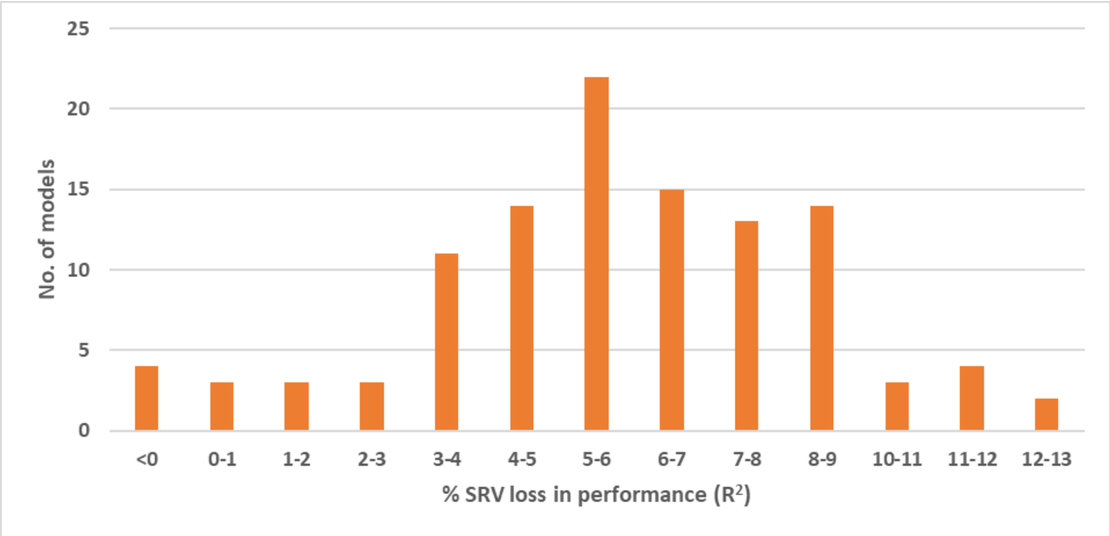


Figure 4.31: SVR models' loss of performance in the prediction of drought severity.

The 111 SVR models posted an average loss in model performance of 6.47. The champion SVR model selected at validation, for example, was recorded to have fallen in performance from an R^2 of 0.83 to 0.78. There, however, was realised a different model from the set of 111 that performed better with an R^2 of 0.83 in the test data set as compared to the SVR champion. This loss of performance between model training and model testing clearly outlines the limitation of the approach of selecting a single best performer model. They vastly lose performance and are therefore unstable for long-term use.

The results of the performance analysis of the SVR ensembles in the prediction of drought severity are as shown in Table 48.

Table 48: Summary performance of the SVR model ensembles in the prediction of drought severity (VCI3M) 1 month ahead

Model	MAE	MAPE	RMSE	R^2
SVR Champion	5.09	0.19	6.95	0.78
SVR Homogeneous Simple Average	4.82	0.18	6.64	0.80
SVR Homogeneous Weighted Average	4.82	0.18	6.65	0.80
SVR Homogeneous Stacked	3.84	0.14	5.09	0.88

The key measure of performance (R^2) for the homogeneous SVR model ensembles realised marginal performance gain of two percentage points between the champion model and the linear and weighted average ensembles. This is as opposed to the use of the ANN techniques in the learning of weights for the stacked ensembling approach that recorded a 10-percentage point improvement in R^2 from an initial R^2 of 0.78 to an R^2 of 0.88. The performance of the SVR homogeneous ensembles as compared to the SVR champion at the county level is provided in Table 49. Each of the counties has the performance provided for the SVR champion and the simple average, weighted average and stacked model ensembles.

Table 49: Performance (R^2) of the SVR homogeneous model ensembles for each county.

Approach	Mandera	Marsabit	Turkana	Wajir	Overall
SVR Champion	0.70	0.77	0.88	0.71	0.78
SVR Homogeneous Simple Average	0.71	0.80	0.87	0.73	0.80
SVR Homogeneous Weighted Average	0.71	0.80	0.87	0.73	0.80
SVR Homogeneous Stacked	0.88	0.85	0.88	0.88	0.88

Except for Turkana county, the homogeneous ensembles outperform the SVR champion model. Turkana, however, records a reduction in performance for the simple average and weighted average approaches. Like the case for homogeneous ANN ensembles, the stacked SVR ensemble guarantees an improvement in the prediction of drought severity at the county level, except for Turkana for which the already good performance is maintained. This performance at the county level is visualized in Figure 4.32.

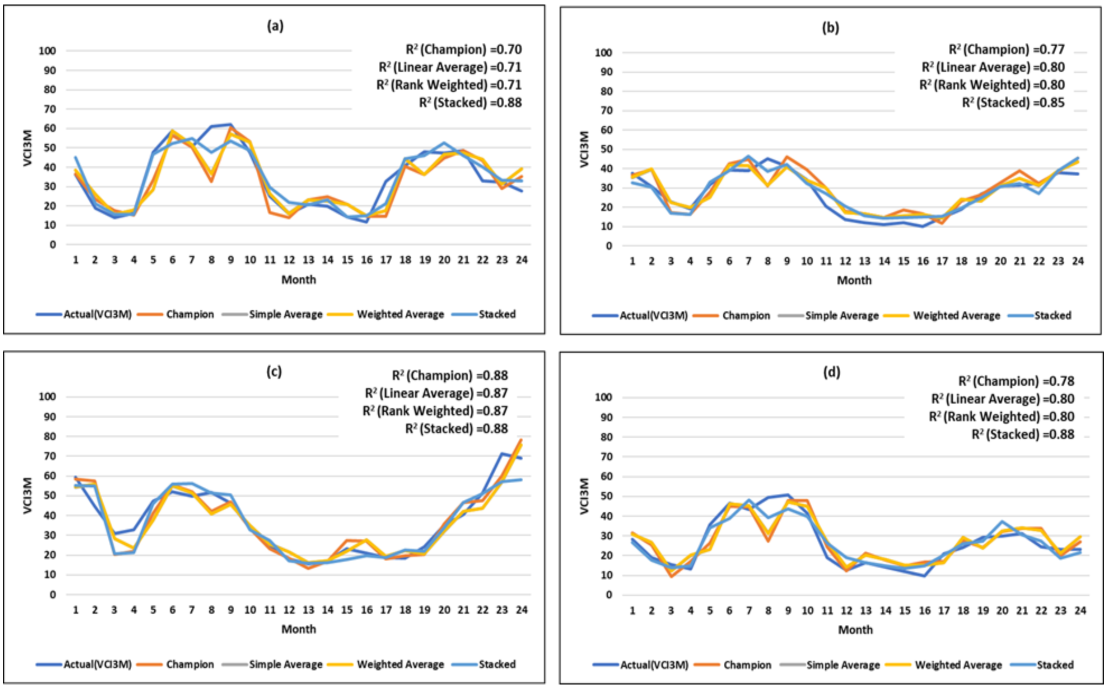


Figure 4.32: Plot of the actual values of VCI3M compared to the predictions from the ANN champion model and the ANN ensemble approaches.

The best performance was registered by the stacked model approach on the SVR models that also offers the closest forecast of the VCI3M values 1 month ahead (Figure 4.32). The trend lines, however, do not provide the final means for judgement due to scale limitations in their interpretation. To augment the analysis of the performance in the prediction of drought severity, analysis of performance in the classification of drought severity was undertaken as presented next.

Classification based performance of the homogeneous SVR ensembles in the prediction of drought severity

The evaluation of the performance of the homogeneous SVR ensembles followed the approach for the homogeneous ANN ensembles. The performance of the homogeneous ensemble classifiers used the key performance measures of- overall model accuracy and Kappa and the class-based performance for model sensitivity and model specificity that are provided in Tables 50-52.

Table 50: Performance by overall accuracy and Kappa for each SVR model ensemble approach

Model	Accuracy	Kappa
SVR Champion	0.69	0.57
SVR Homogeneous Simple Average	0.69	0.57
SVR Homogeneous Weighted Average	0.70	0.58
SVR Homogeneous Stacked	0.77	0.69

Table 51: Performance by Sensitivity for each SVR model ensemble approach

Model	Class				
	1	2	3	4	5
SVR Champion	-	0.77	0.64	0.69	0.70
SVR Homogeneous Simple Average	NA	0.86	0.62	0.65	0.70
SVR Homogeneous Weighted Average	NA	0.86	0.62	0.65	0.70
SVR Homogeneous Stacked	NA	0.86	0.63	0.68	0.70

Table 52: Performance by Specificity for each SVR model ensemble approach

Model	Class				
	1	2	3	4	5
SVR Champion	0.99	0.89	0.84	0.90	0.95
SVR Homogeneous Simple Average	0.99	0.87	0.88	0.89	0.95
SVR Homogeneous Weighted Average	0.99	0.87	0.89	0.89	0.95
SVR Homogeneous Stacked	0.99	0.91	0.90	0.91	0.98

The Kappa value for the SVR champion, the linear average and the weighted average all recorded good agreement with that of the stacked model.

The averaging approaches to the creation of SVR model ensembles, both simple average and weighted average approaches, produced models that offered marginal improvement in classification accuracy as compared to the SVR champion model. Like the case for homogeneous ANN ensembles, the SVR stacked model ensemble realised the best return in overall model performance. The analysis of the performance of the homogeneous SVR ensembles at county level returned the results in Table 53.

Table 53: Classification accuracy for the SVR homogeneous ensembles.

Approach	Mandera	Marsabit	Turkana	Wajir	Overall
SVR Champion	0.58	0.75	0.83	0.58	0.69
SVR Homogeneous Simple Average	0.63	0.83	0.67	0.63	0.69
SVR Homogeneous Weighted Average	0.63	0.83	0.71	0.63	0.70
SVR Homogeneous Stacked	0.79	0.88	0.75	0.71	0.78

The simple average homogeneous SVR ensemble is shown in Table 53 to have offered no overall improvement in performance. In fact, for the case of Turkana, it massively loses performance in the classification of drought severity. The best overall performance is realised from the SVR stacked ensemble that also posted the lowest range in performance of 0.17 as compared to the champion model that has the highest range of 0.25, therefore, making for a more unbalanced performance in the prediction of drought severity 1 month ahead.

The results of the calculation of the area under the ROC (AUROC) for the SVR ensemble approaches in Table 54 clearly shows the SVR stacked model ensemble as posting an AUROC of 91.2%.

Table 54: Classification accuracy for the SVR homogeneous ensembles.

Approach	AUROC
SVR Champion	88.54
SVR Homogeneous Simple Average	90.29
SVR Homogeneous Weighted Average	90.44
SVR Homogeneous Stacked	91.20

As compared to the other approaches, the model stacking approach to SVR homogeneous model ensembles offer the best return in investment from the computational requirements as compared to say model weighted averaging that just marginally improves on the performance of the simple averaging approach.

4.3.2.3 Homogeneous ANN model ensembles in the prediction drought effects

The study built 976 models for the prediction of drought effects. A total of 272 models of both ANN and SVR were selected for model ensembling based on a cut off of $R^2 \geq 0.7$ and further ensemble size reduction following the backward elimination with forward substitution approach. The investigation of the performance of the model ensembles followed on the three ensemble approaches as is the case in the prediction of drought severity. We investigated the performance of the homogeneous ANN ensembles in both regression and in the prediction of drought classes.

Regression based performance of the homogeneous ANN ensembles in the prediction of drought effects

The performance of the ANN ensembles on the prediction of drought effects is summarized in Table 55 using the R^2 and 3 other error-based measurements of performance on the test dataset.

Table 55: Summary performance of the homogeneous ANN ensembles in the prediction of drought effects (MUAC) 1 month ahead.

Model	MAE	MAPE	RMSE	R^2
ANN Champion	1.49	0.08	2.13	0.74
ANN Homogeneous Simple Average	1.53	0.08	2.07	0.75
ANN Homogeneous Weighted Average	1.42	0.08	1.95	0.77
ANN Homogeneous Stacked	1.28	0.07	1.74	0.82

The best ANN model for the prediction of drought effects which is treated as the base performance had an R^2 of 0.74 on the overall test data set covering the period 2016-2017. The performance of the ANN models in the prediction of drought effects indicates the superiority of the stacking approach in the creation of model ensembles as compared to the simple and weighted averaging approaches. In fact, simple averaging is shown not to have offered considerable improvement in performance given it posted only a 1 percentage point increase over the champion model. The gain in performance of the model stacking approach of 8 percentage points is in our opinion worth the investment in the complexity of the process. This judgement is also supported by the RMSE that also indicates the stacked model ensembling as posting the lowest and hence best variance in the prediction of drought effects.

The performance of the ANN homogeneous ensembles in the prediction of drought effects disaggregated and compared to that of the ANN champion at the county level for each of the model ensembling approaches is provided in Figure 4.33.

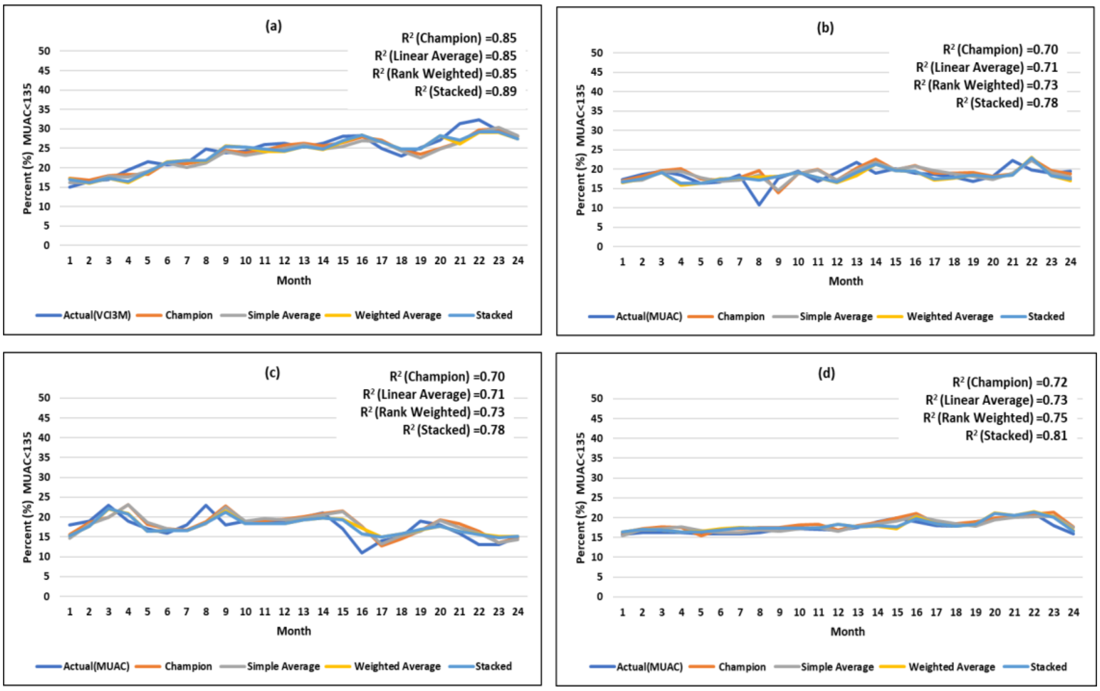


Figure 4.33: Plot of the actual values of children under the risk of malnutrition compared to the predictions from the ANN champion model and the ANN ensemble approaches.

Figure 4.33 shows the performance plotted for each of the ensemble approaches of linear averaging, rank-weighted averaging and stacking as compared to the champion model. The best performance was posted by the stacked ensemble with an overall R^2 of 0.82. The homogeneous ANN stacked ensemble thus offers the best improvement over the ANN champion in the prediction of drought effects.

Classification based performance of the homogeneous ANN ensembles in the prediction of drought effects

The aggregation of proportions of children at risk of malnutrition to form the nutrition status of an administrative unit and its subsequent thresholding into classes is not documented. Before the actual thresholding, this study investigated normality for the proportion of children with MUAC<135 in order to inform the method of analysis to be deployed. The histogram for MUAC is as shown in Figure 4.34.

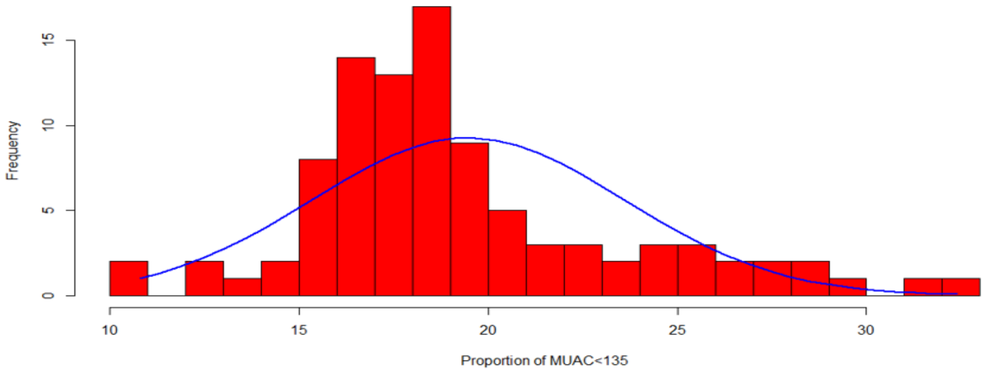


Figure 4.34: Histogram plots for the proportion of children with MUAC<135

On visual inspection of the histogram plot, the proportion of MUAC<135 amongst the children below 5 years is realised to be non-normally distributed. This is contrary to the reliance of many studies on the normality of the distribution of proportional of children under the risk of malnutrition (MUAC<135) in nutrition surveys. Statistical testing for normality using both Shapiro-Wilk and Anderson-Darling test for normality were done with the results provided in Table 56.

Table 56: Normality testing for the proportion of children under 5 years with MUAC<135

Test	Statistic name	Statistic value	P-value	Decision on H_o
Shapiro-Wilk	Shapiro-Wilk	0.95	0.001	Reject
Anderson-Darling	Anderson-Darling	2.94	1.88e-07	Reject

The Null hypothesis (H_o) for either test is that "the data is normal". In both cases, the p-values, p are less than 0.05 at the 95% confidence level. The null hypotheses in both cases were rejected and thus the data are concluded to be non-normally distributed. This confirms the earlier conclusion from the visual inspection of the plot.

With the proportion of children under threat of malnutrition established as non-normally distributed, the definition and analysis of the classes on the drought effects, therefore, followed a non-parametric approach as opposed to being based on the standardization of the MUAC values. This is akin to the calculation of MUAC condition index (MCI). The classes were then defined as shown in Equation 35 with the distribution of the classes subsequently shown in Figure 4.35. The higher the proportion of MUAC<135, the worse the measure for nutrition is. 1 implies drought affects MUAC while 0 implies MUAC not affected. Scaling was done for each county separately.

$$MUACstatus = \begin{cases} 1, if\ 100 - 100 * MCI < 50 \\ 0, otherwise \end{cases} \quad \dots (35)$$

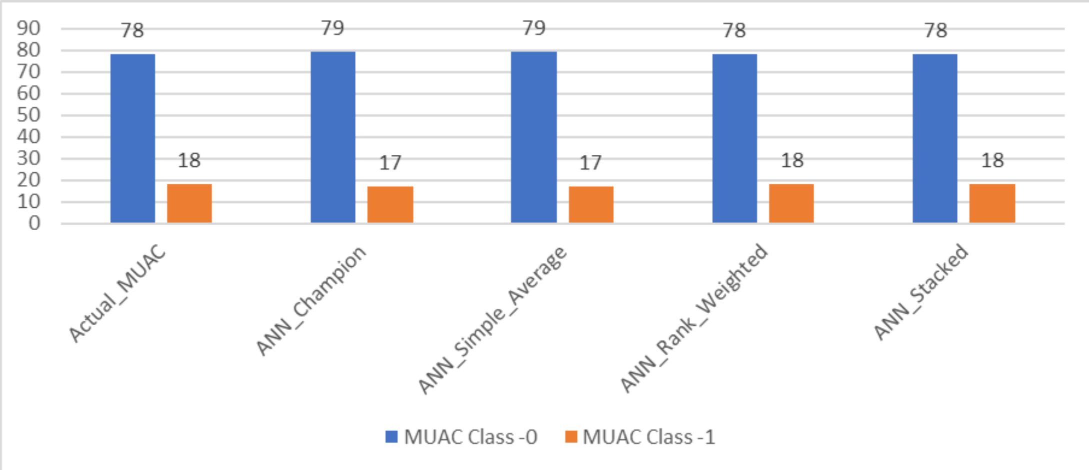


Figure 4.35: The distribution of actual MUAC classes and their prediction from the ANN champion and the ensemble approaches.

Despite the minor differences in the proportion of the instances classified as either drought-affected by MUAC (class 1, orange line), there exist differences in the performance of the classifier as shown by the Accuracy, Kappa, and AUROC provided in Table 57.

Table 57: Performance of the ANN-based classification of drought effects

Model	Accuracy	Kappa	Sensitivity	Specificity	AUROC
ANN Champion	0.95	0.83	0.96	0.88	90.38
ANN Homogeneous Simple Average	0.95	0.83	0.96	0.88	90.38
ANN Homogeneous Weighted Average	0.96	0.86	0.97	0.89	93.16
ANN Homogeneous Stacked	0.96	0.86	0.97	0.89	93.16

Evidently, as a binary classifier, the ANN ensembles remain competitive compared to the champion model. It is, however, the case that the gain in performance is marginal using the measures of accuracy, kappa, sensitivity and specificity. An investigation of the sensitivity- specificity trade-off (AUROC) however reveals the superiority of the ranked weighted and stacked model ensembles as compared to the other approaches.

All the ANN approaches produced very good agreements with Kappa ≥0.8 together with very high accuracies.

4.3.2.4 Homogeneous SVR model ensembles in the prediction drought effects

Regression based performance of the homogeneous SVR ensembles in the prediction of drought effects

Similar to the homogeneous ANN model ensembles, the SVR ensembles had a total of 272 SVR models chosen from the earlier set of 976. It is for these 272 models that the results of model ensembling were based. Table 58 presents the summary of the performance of the SVR champion and the SVR ensemble models based on the different ensembling approaches.

Table 58: Performance of the SVR ensemble models in the prediction of drought effects

Model	MAE	MAPE	RMSE	R^2
SVR Champion	1.62	0.09	2.25	0.71
SVR Homogeneous Simple Average	1.60	0.09	2.21	0.72
SVR Homogeneous Weighted Average	1.58	0.09	2.19	0.72
SVR Homogeneous Stacked	1.44	0.08	1.99	0.77

The best model performance as indicated by R^2 is registered by the stacked modelling approach that offers at least a 5 percentage points gain in performance as compared to the other approaches. Compared to the champion model approach, the gain in performance is a considerable 6 percentage points. The measures of error are inverse in relationship to the R^2 values with the stacking approach producing the lowest measures of error.

Overall, as presented in Table 58, both the simple and weighted averaging approaches offered only a 1 percentage point improvement in performance and would be considered not of practical advantage compared to the computational and handcrafting complexities that accompany them.

A visualization of the performance of the homogenous ensembling approaches on the test dataset is provided in Figure 4.36.

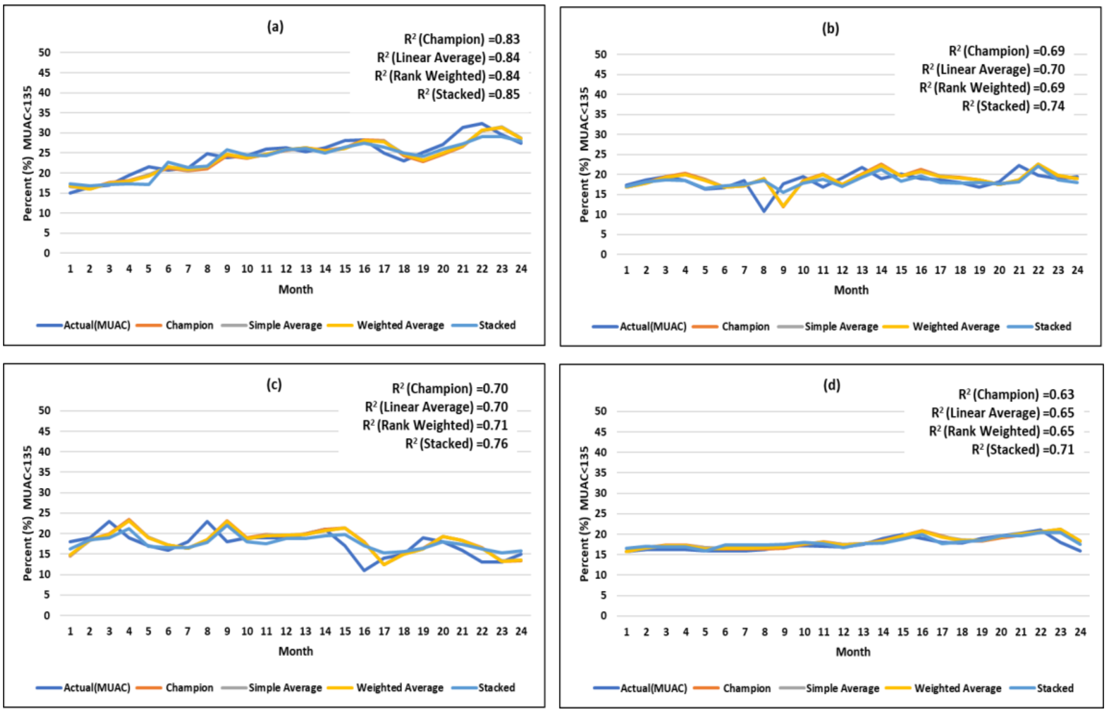


Figure 4.36: Plot of the actual values of children under the risk of malnutrition compared to the predictions from the SVR champion model and the SVR ensemble approaches.

Like in the earlier approaches, visual inspection of Figure 4.36 together with the models' R^2 indicates the closest relationship as that between the predictions from the stacked model ensemble and the actual MUAC proportions in comparison to the other model ensemble approaches.

Classification based performance of the homogeneous SRV ensembles in the prediction of drought effects

Following on the binary definition of the MUAC classification problem from the ANN homogeneous ensemble models, the distribution of the classes in the SVR approach is as depicted in Figure 4.37.

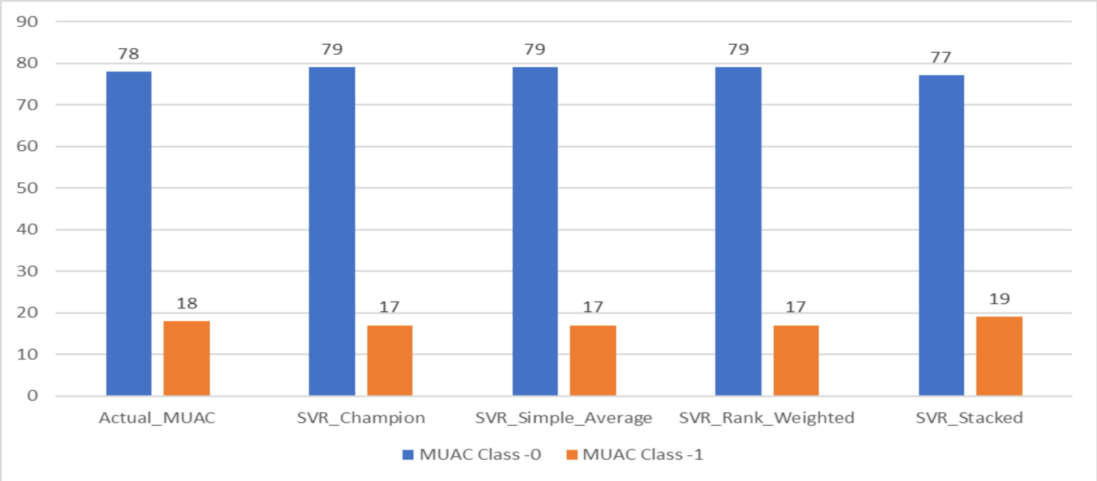


Figure 4.37: The distribution of actual MUAC classes and their prediction from the SVR champion and the ensemble approaches.

The distribution of the classes within each model ensembling approach is fairly equal in proportion at between 17 and 19. However, this position changes when the actual month on month performance of the classifiers is investigated as shown in Table 59.

Table 59: Performance of the SVR ensemble models in the classification of drought effects

Model	Accuracy	Kappa	Sensitivity	Specificity	AUROC
SVR Champion	0.95	0.83	0.96	0.88	90.38
SVR Homogeneous Simple Average	0.95	0.83	0.96	0.88	90.38
SVR Homogeneous Weighted Average	0.95	0.83	0.96	0.88	90.38
SVR Homogeneous Stacked	0.95	0.83	0.97	0.84	92.52

In terms of accuracy, kappa, sensitivity and specificity as stand-alone measures of model performance, all the model ensemble approaches almost post the same performances. The simple and weighted averaging approaches post the same performance across the metrics while the stacking approach is indicated to have produced SVR models that were marginally better in model sensitivity by a percentage point. Further analysis on the trade-off between sensitivity and specificity and a calculation of area under the ROC (AUROC) however settled on the superiority of the model stacked ensemble with an AUROC of 92.52 and thus an improvement of over 2 percentage points in comparison to the other approaches that each had an AUROC of 90.38.

4.3.2.5 Summary on Building and evaluation of homogeneous model ensembles

To answer to the sixth research question (RQ6) on **"***What is the performance of the Artificial Neural Networks (ANN) and Support Vector Regression (SVR) homogeneous ensemble models in the prediction of both drought severity and drought effects?"*, we used the ANN and SVR machine learning techniques of to develop model ensembles following on the three model ensembling approaches of simple averaging, weighted averaging and model stacking.

The performances of the model ensembles were evaluated against reality using an out of sample testing data set of 24 data points covering the years 2016-2017 for each of the four counties in the study area. Different metrics were used to evaluate the model performance including R^2 as the primary performance metrics. Mean absolute error (MAE), mean absolute percentage error (MAPE) and root mean squared error (RMSE) were also calculated. For the prediction of drought severity formulated as a classification problem, the metrics used were accuracy, Kappa, sensitivity and specificity. A final investigation of performance with a trade-off of sensitivity to specificity was done for each of the prediction of drought severity and drought effects using the AUROC.

The investigation of normality of the proxy measure of drought effects (MUAC) returned non-normality. The definition of the classes for MUAC, therefore, was done using a scaling following on relative range difference. A binary classification was then defined over the drought effects variable as opposed to drought severity that is defined over a set of 5 classes.

For drought severity, the performance of the approaches using a summary of key metrics is as presented in Table 60. The metrics include R^2, overall accuracy and AUROC.

Table 60: Summary of performance of ensemble models in drought severity prediction

	R^2		Accuracy		AUROC	
Approach	ANN	SVR	ANN	SVR	ANN	SVR
Champion	0.82	0.78	0.71	0.69	89.37	88.54
Homogeneous Simple Average	0.84	0.80	0.70	0.69	90.83	90.29
Homogeneous Weighted Average	0.85	0.80	0.69	0.70	90.63	90.44
Homogeneous Stacked	0.91	0.88	0.78	0.77	91.26	91.20

The stacking approach produced better predictions across both homogeneous ANN and homogeneous SVR model ensembles as indicated by both R^2 and AUROC for these model ensembles as compared to the ANN and SVR champion models. The improvement in performance derived from the linear and weighted average approaches range from a marginal 2 percentage points for the SVR technique to 3 percentage points for the ANN technique. As indicated earlier by Kappa, the better performing ensembles have a higher probability of having been arrived at by chance. Table 61 provides a summary of the key performance metrics in the prediction of drought effects similar to those provided for the prediction of drought severity.

Table 61: Summary of performance of ensemble models in drought effects prediction

Approach	**R^2**		**Accuracy**		**AUROC**	
	ANN	**SVR**	**ANN**	**SVR**	**ANN**	**SVR**
Champion	0.74	0.71	0.95	0.95	90.38	90.38
Homogeneous Simple Average	0.75	0.72	0.95	0.95	90.38	90.38
Homogeneous Weighted Average	0.77	0.72	0.96	0.95	93.16	90.38
Homogeneous Stacked	0.82	0.77	0.96	0.95	93.26	92.52

The performance of the homogeneous ensembles in the prediction of drought effects mirrors that of drought severity. The performance of the simple average and weighted average homogeneous model ensembles are just marginally above those of the champion model as measured by R^2, accuracy and AUROC. In general, the ANN ensembles also out-performed the SVR ensembles for each ensembling approach by up to 5 percentage points of R^2. As classifiers, these differences become virtually lost at almost zero percentage point difference especially in the case of simple averaging.

The answer to the question on the performance of the homogeneous ensembles in the prediction of drought severity and the prediction of drought effects we documented the superiority of any homogeneous model ensemble over any of the non-ensembling approaches. Model stacking is guaranteed to produce the most predictive models as compared to non-ensembled champion models. The gain in performance by stacking is up to a massive 10 percentage points above the performance of the champion models. The achievement of an R^2 of 0.91 in the prediction of drought severity is in fact considered an excellent model performance in the prediction of drought conditions 1 month ahead. For the prediction of drought effects, the improvement in performance as a result of stacking as an approach to model ensembling is 8 percentage points for ANN and 6 percentage points for SVR over the respective ANN champion and SVR champion models respectively. These good evaluation results based on R^2 for the stacked ensemble were supported by the AUROC as a measure of performance in classification that saw an improvement of 1.89 percentage points for the ANN technique and 2.66 percentage points for the SVR technique in the prediction of drought severity. This was as compared to 2.88 percentage points for the ANN champion and 2.14 percentage points for the SVR champion in the prediction of drought effects.

4.3.3 Heterogeneous model ensembles

The seventh research question (RQ7) was formulated as **"***What is the performance of the ANN and SVR heterogeneous ensemble models in the prediction of drought severity and drought effects?".* To answer this research question (RQ7), we built heterogeneous ensemble models from the same number of models earlier chosen for model ensembling. The models for ensembling were: 111 models of both ANN and SVR for the prediction of drought severity (VCI3M) 1 month ahead and 272 models of both ANN and SVR for the prediction of drought effects (MUAC) 1 month ahead. The heterogeneous ensembles, therefore, have 222 models and 544 models for the prediction of drought severity and drought effects respectively.

Heterogeneous model ensembles in the context of this study are ensembles in which the models used in the ensemble are built using more than one machine learning technique- the ANN and SVR techniques in the context of this study. The performance of the heterogeneous model ensembles were evaluated against the performance of the champion models from the ANN and SVR techniques: the ANN champion and SVR champion respectively.

4.3.3.1 Heterogeneous model ensembles for the prediction of drought severity

The model selection stage of the model ensembling process presented in Figure 4.25 ended up with 111 models of ANN and 111 models of SVR for model ensembling. The use of both sets of ANN and SVR models in prediction is called heterogeneous ensembling.

To realize comparable results, we used the same 111 ANN and 111 SVR models selected from the ensembles membership composition stage to develop the heterogeneous model ensembles. The models were ensembled using the three approaches: of simple averaging, weighted averaging and model stacking approach also referred to as meta-model averaging in which we used an ANN meta-model to learn the model weights. The ensemble realized from the averaging approaches was thus referred to as Heterogeneous Simple Average and Heterogeneous Weighted Average while the ensemble from stacking was referred to as the Heterogeneous

Stacked ensemble. In the weighted approach, each model was weighted based on its performance in the validation dataset before judging the model's contribution to prediction in the test dataset.

Regression based performance of the heterogeneous model ensembles in the prediction of drought severity

From the previous results, it was established that the best performing champion models in the prediction of drought severity in the out of sample test dataset recorded an R^2 of 0.82 and 0.78 respectively for the ANN and SVR techniques. The loss of performance of the SVR modelling process of 4 percentage points in the test dataset was a good illustration of the need for model ensembles over single champion models. This loss of performance defines the lack of stability in the performance of single champion models.

The choice of the model to be used as the base model for comparison was not a naïve decision. Three options were available for the formulation of the performance of the base model. The first was to assume the minimum performance of R^2 of 0.78 as was posted by the SVR champion. This would have biased against the base models given the choice of the poorest performing technique. The second option was the averaging of the performance of the base models to an R^2 of 0.80 for the ANN and SVR champions that had posted an R^2 of 0.82 for the ANN and an R^2 of 0.78 respectively. The third option was to assume the model outputs as representing prediction of the same data points in space. The two predictions were then averaged and new performance evaluation was undertaken. This raised an R^2 of 0.82. The latter two options were judged to be ensemble approaches in themselves and would not amount to a pure selection of champions as used in most modelling approaches. We, therefore, settled on the comparison with both the ANN champion and the SVR champion model's performance. The base R^2 was therefore set at 0.78 for the SVR champion and at 0.82 for the ANN champion. The desired performance is an R^2 of greater than 0.82 for the heterogeneous ensembles. More desirable was a heterogeneous ensemble that outperforms both the ANN and SVR champion models while also outperforming the homogeneous ensembles.

The performance of the heterogeneous model ensemble in the prediction of drought severity as compared to the ANN champion and the SVR champion is provided in Table 62.

Table 62: Summary performance of the heterogeneous model ensemble in the prediction of drought severity (VCI3M) 1 month ahead

	MAE	MAPE	RMSE	R^2
ANN Champion	4.74	0.18	6.31	0.82
SVR Champion	5.09	0.19	6.95	0.78
Heterogeneous Simple Average	4.68	0.18	6.39	0.82
Heterogeneous Weighted Average	4.56	0.17	6.22	0.82
Heterogeneous Stacked	2.81	0.11	3.77	0.94

The heterogeneous ensembles have a mixed set of performances in the prediction of drought severity as compared to the ANN and SVR champion models. All the heterogeneous ensembles out-perform the SVR champion. The ANN champion, however, remains competitive, at an R^2 of 0.82, to the heterogeneous ensembles realised from the averaging approaches of both linear and weighted average ensembles. Evidently, the stacked heterogeneous model ensemble posts a distinct improvement in the prediction of drought severity as indicated by VCI3M value 1 month ahead at an R^2 of 0.94. The stacked model approach to ensembling is, in this case, a distinctly superior method in the prediction of future drought conditions as compared to the selection of single champion models.

An extended analysis of the performance of the heterogeneous model ensemble in the prediction of drought severity posed as a regression problem is provided in Figure 4.38. The extended analysis was for each of the 24 data points in the out-of-sample datasets covering the period 2016-2017 for each of the four counties in the study area. This makes for a total of 96 data points for the study area.

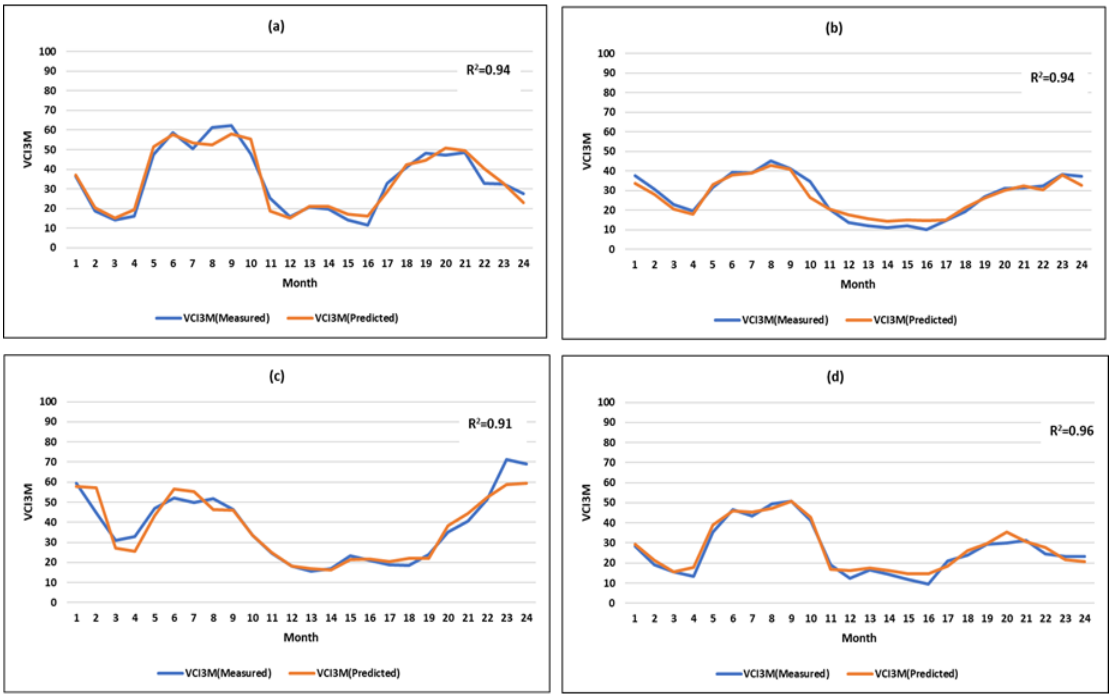


Figure 4.38. A plot of the actual values of drought severity (VCI3M) versus the values predicted 1-month ahead from the heterogeneous stacked ensemble.

The heterogeneous stacked ensemble realised a performance for: (a) Mandera (R^2=0.94); (b) Marsabit (R^2=0.94); (c) Turkana (R^2=0.91) and (d) Wajir (R^2=0.96). The performance in the prediction of VCI3M values 1 month ahead, therefore, realized an excellent prediction with R^2 between 0.91 and 0.96 across all the counties. The performance across the counties posted an acceptable distribution at the county level. The performance at the county level is particularly important given it is for a model developed for a bigger administrative unit but whose performance at prediction is analysed to be acceptable at downscaled lower administrative units.

Classification based performance of the heterogeneous model ensembles in the prediction of drought severity.

The presentation of the results of the performance of the heterogeneous model ensembles as a classification problem follows on the formulation of drought effects classes on the proxy variable MUAC as was the case for homogeneous models. Tables 63 to 65 document the performance of the heterogeneous model ensembles when formulated as a classifier.

Table 63: Performance of the heterogeneous model ensemble by overall accuracy and Kappa in the prediction of drought severity

Model	**Accuracy**	**Kappa**	**AUROC**
ANN Champion	0.71	0.60	89.37
SVR Champion	0.69	0.57	88.54
Heterogeneous Simple Average	0.70	0.58	90.44
Heterogeneous Weighted Average	0.71	0.60	90.45
Heterogeneous Stacked	0.80	0.73	92.26

Table 64: Performance of the heterogeneous model ensemble by sensitivity in the prediction of drought severity

Model	Class				
	1	2	3	4	5
ANN Champion	0	0.82	0.63	0.73	0.78
SVR Champion	0	0.77	0.64	0.69	0.70
Heterogeneous Simple Average	NA	0.86	0.63	0.68	0.70
Heterogeneous Weighted Average	NA	0.86	0.64	0.68	0.70
Heterogeneous Stacked	NA	0.88	0.77	0.86	0.67

Table 65: Performance of the heterogeneous model ensemble by specificity in the prediction of drought severity

Model	Class				
	1	2	3	4	5
ANN Champion	0.99	0.86	0.88	0.91	0.95
SVR Champion	0.99	0.89	0.84	0.90	0.95
Heterogeneous Simple Average	0.99	0.87	0.89	0.89	0.95
Heterogeneous Weighted Average	0.99	0.88	0.89	0.89	0.95
Heterogeneous Stacked	0.99	0.92	0.93	0.91	0.99

The heterogeneous simple average ensemble in Table 63 performs poorer than the ANN champion model in overall model accuracy. It is also noted that the better the model performance the higher the Kappa thereby implying the higher the probability that the resultant classifier models were obtained by chance.

In terms of sensitivity as provided in Table 64, the first two classes marked by low vegetation cover are noted to easier return true positive predictions as compared to the other classes. This performance is replicated in specificity as shown in Table 65 where the classes 2-4 have lower performance across all the ensembling approaches as compared to classes 1 and 5. This implies that the heterogeneous models are efficient at predicting the outlier drought conditions as compared to the other classes of drought.

An investigation into the trade-off between sensitivity and specificity of the heterogeneous models for the prediction of drought sensitivity, however, realised an area under the ROC (AUROC) of 92.26% for the heterogeneous stacked model ensemble as indicated in Table 63. The heterogeneous stacked ensemble is thus noted to out-perform the other approaches to model ensembling and is trailed by both the linear averaged and weighted averaged ensembles that post almost a similar performance with an AUROC of 90.44% and 90.45% respectively.

Further analysis of the performance of the heterogeneous model ensembles in the prediction of drought severity classes at the county level using model accuracy is provided in Table 66.

Table 66: Classification accuracy of the heterogeneous ensemble

	Mandera	**Marsabit**	**Turkana**	**Wajir**	**Overall**
ANN Champion	71	75	71	67	71
SVR Champion	58	75	83	58	69
Heterogeneous Simple Average	63	83	71	63	70
Heterogeneous Weighted Average	63	83	71	67	71
Heterogeneous Stacked	**71**	**88**	**79**	**83**	**80**

It is quite evident from Table 66 that the model stacking approach to model ensembling offers quite a good return in predictive accuracy at 80% as compared to the linear average and the weighted approaches that are by and large only comparable in performance with each other and with the best performing champion model- the ANN champion model. At the county level, all the counties registered an accuracy of more than 70% with the best at 88% in Marsabit. A visualization of the month on month performance of the heterogenous stacked classifier at the county level is provided in Figure 4.39.

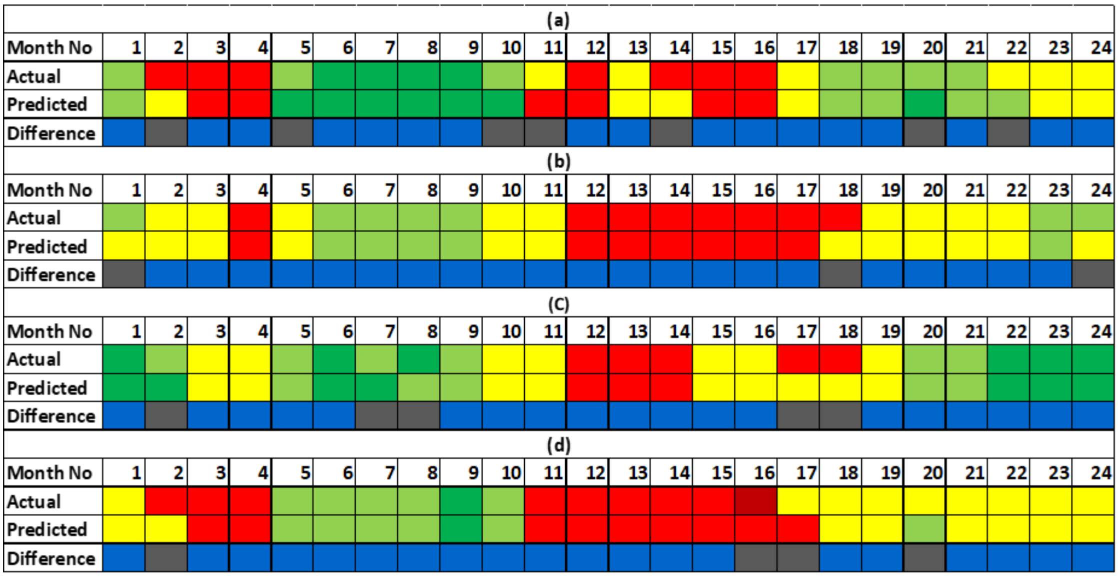


Figure 4.39. Performance of the heterogeneous ensemble classifier for each of the counties.

The months of difference in Figure 4.39 are presented in grey while those of agreement are in blue. The performance of the heterogeneous stacked classifiers with an overall accuracy of 80% is superior to that of the best champion model, the ANN champion model, that posted an accuracy of 71% over the entire test data set. Even at the county level, the heterogeneous stacked classifier outperforms the champion classifiers across all counties except for Turkana where the SVR champion performs better as earlier highlighted in Table 66 in **bold**. It is this superiority of the SRV model in the Turkana dataset that perhaps lends the opportunity for model ensembling.

Give the interest of this study in the correctness of the prediction of moderate to extreme drought, a final analysis was done on the utility of the heterogeneous stacked ensemble in the prediction of moderate to extreme drought classes. This view of analysis was informed by the pilot study that earlier showed the poor performance of the champion models in the prediction of moderate to extreme drought classes in the prediction of drought severity as was presented in Figure 3.32. Table 67 presents the performance of the stacked ensembles in the prediction of moderate to extreme vegetation deficits across the different classes for the counties in the study area as compared to the champion models.

Table 67: Model accuracy in the prediction of moderate to extreme drought of the heterogeneous stacked ensembles compared to the champion ANN and SVR models

County	ANN Champion	SVR Champion	Heterogeneous Stacked Ensemble
Mandera	62	46	69
Marsabit	71	71	94
Turkana	75	100	83
Wajir	72	61	78
Overall	70	69	82

The heterogeneous stacked ensemble is shown in Table 67 to offer the best performance in the prediction of moderate to extreme vegetation deficit. The distributions across the counties are also acceptable as compared, for example, to the SVR champion that performs below chance with an accuracy less than 50% for Mandera county. Two things particularly stood out in the performance of the heterogeneous ensemble:

- The relatively high predictive power of the heterogeneous ensemble at an average accuracy of 82% as compared to 70% and 69% for the ANN champion and SVR champion respectively.
- The relatively moderate range between the best and the worst performance at the county level with a difference of 25 percentage points difference as compared to say the SVR champion that posted a difference of 54 percentage points.

4.3.3.2 Heterogeneous model ensembles for the prediction of drought effects

The ANN and SVR champion models in the prediction of drought effects had an R^2 of 0.74 and 0.71 respectively. Like in the homogeneous ensembles, the prediction of drought effects using heterogeneous ensembles was done using the top 272 models of both ANN and SVR.

Regression based performance of heterogeneous model ensembles in the prediction of drought effects

Table 68 presents the performance of the heterogeneous model ensembles of ANN and SVR in the prediction of drought effects following on the different model ensembling approaches.

Table 68: Performance of the heterogeneous model ensembles in the prediction of drought effects (MUAC)

	MAE	**MAPE**	**RMSE**	**R^2**
ANN Champion	1.49	0.08	2.13	0.74
SVR Champion	1.62	0.09	2.25	0.71
Heterogeneous Simple Average	1.54	0.08	2.09	0.75
Heterogeneous Weighted Average	1.65	0.09	2.15	0.78
Heterogeneous Stacked	1.41	0.08	1.84	0.82

The stacked heterogeneous ensemble approach out-performs the simple average and ranked weighted ensemble approaches in all the metrics of performance measurement. For the prediction of drought effects, the stacking approach offers an 8 and 11 percentage points improvement in performance as compared to the ANN champion and the SVR champion models respectively. All ensembling approaches also perform better than the use of the champion models.

Classification based performance of heterogeneous model ensembles for drought effects

Using the ensemble models as a drought effects classifier produces the results presented in Table 69.

Table 69: Performance of the heterogeneous ensemble in the classification of drought effects

Model	Accuracy	Kappa	Sensitivity	Specificity	AUROC
ANN Champion	0.95	0.83	0.96	0.88	90.38
SVR Champion	0.95	0.83	0.96	0.88	90.38
Heterogeneous Simple Average	0.95	0.83	0.96	0.88	90.38
Heterogeneous Weighted Average	0.93	0.74	0.94	0.87	84.83
Heterogeneous Stacked	0.96	0.86	0.97	0.89	93.16

The highest accuracy in the classification of drought effects is realised from the model stacking approach. In the classification of drought effects, all the ensembling approaches except weighted average remain competitive based on the different performance measures of accuracy, kappa, sensitivity and specificity. The AUROC clearly confirms the poor performance of the weighted ensemble approach to model ensembling. The weighted ensembles posted a performance that falls below the performance of the ANN champion and SVR champion in the prediction of MUAC conditions 1 month ahead by up to 5.5 percentage points.

4.3.3.3 Summary on Building and evaluation of heterogeneous model ensembles

To answer to the research question **RQ7 "***What is the performance of the ANN and SVR heterogeneous ensemble models in the prediction of drought severity and drought effects?"* we built heterogeneous ensemble models of both ANN and SVR using the three different approaches of simple averaging, weighted averaging and ANN-driven model ensembling.

In the prediction of drought severity, a summary of the performance of the different heterogeneous ensembles is presented in Table 70. The summary compares the performance of the heterogeneous model ensembles to those of the homogeneous ensembling approaches. The base models are the champion modes for the prediction of drought severity with an R^2 of 0.82 and 0.78 for the ANN and the SVR techniques respectively.

Table 70: Performance of heterogeneous as compared to homogeneous ensembles in the prediction of drought severity (VCI3M)

Approach	R^2	RMSE	Accuracy	Kappa	AUROC
Champion ANN	0.82	6.31	0.71	0.60	89.37
Champion SVR	0.78	6.95	0.69	0.57	88.54
Homogeneous Simple Average (ANN)	0.84	5.96	0.70	0.58	90.83
Homogeneous Simple Average (SVR)	0.80	6.64	0.69	0.57	90.29
Heterogeneous Simple Average	0.82	6.39	0.70	0.58	90.44
Homogeneous Weighted Average (ANN)	0.85	5.85	0.69	0.57	90.63
Homogeneous Weighted Average (SVR)	0.80	6.65	0.70	0.58	90.44
Heterogeneous Weighted Average	0.82	6.22	0.71	0.60	90.45
Homogeneous Stacked (ANN)	0.91	4.40	0.78	0.70	91.26
Homogeneous Stacked (SVR)	0.88	5.09	0.77	0.69	91.20
Heterogeneous Stacked	0.94	3.77	0.80	0.73	92.26

The comparison between heterogeneous and homogeneous model ensembles returns mixed results. The ensembles remain competitive in their performance between the homogeneous and the heterogeneous and as compared to the champion models. Cases of loss of performance were recorded for the weighted ensembles. It is, however, a guarantee that using model stacking on heterogeneous model ensembles posts better results than any of the ensembling approaches both for homogeneous ensembles and for heterogeneous ensembles. Stacked heterogeneous model ensembles registered the best performance in both classification and regression in the prediction of drought severity.

The performance of the ensembles as compared to the champion models in the prediction of drought effects on malnutrition (MUAC) is as provided in Table 71.

Table 71: Summary of performance in prediction of drought effects (MUAC)

Approach	Regression ($R^{2)}$	Accuracy (%)
Champion ANN	0.74	95
Champion SVR	0.71	95
Homogeneous Simple Average (ANN)	0.75	95
Homogeneous Simple Average (SVR)	0.72	95
Heterogeneous Simple Average	0.75	95
Homogeneous Weighted Average (ANN)	0.77	96
Homogeneous Weighted Average (SVR)	0.72	95
Heterogeneous Weighted Average	0.78	93
Homogeneous Stacked (ANN)	0.82	96
Homogeneous Stacked (SVR)	0.77	95
Heterogeneous Stacked	0.82	96

In the prediction of drought effects, the performance of the heterogeneous model ensembles reinforces the superiority of stacking approach to model ensembling with an R^2 of 0.82 and an accuracy of 96%.

4.4 Highlight of Key Findings in Context

In this sub-section, we highlight the key findings of the study in the context of related studies. Higher correlations were recorded between the vegetation datasets and drought severity as compared to the precipitation datasets and the other datasets. This is an expected consequence of the definition of drought severity in terms of vegetation deficit that therefore makes variables on vegetation conditions good predictors. On the other hand, rainfall anomalies may or may not lead to significant changes in vegetation conditions as this could be a function of available water resources. It is for this reason that many studies on agricultural drought prefer observing directly the vegetation conditions, without relying on rainfall data that are often sparse for actual measurements and suffer high uncertainties in the case of modelled outputs (Dinku et al., 2007).

It is the expectation in theory that heterogeneous ensembles should outperform both homogeneous and champion model ensembles in both regression and classification. This is widely the case in this study that mirrors similar results documented in Petrakova, Affenzeller & Merkurjeva (2015). With an overall R^2 of 0.94 in regression and an accuracy of 80% in classification, the stacked heterogeneous model ensemble is superior to both the homogeneous ensembles and the champion model approach to the prediction of future vegetation conditions as a proxy in the prediction of drought severity. The superiority of the ensembles against champion models especially, as indicated by overall prediction accuracy, is however not guaranteed for generalization across the spatial units. The case in Table 66 has the ANN champion and the SVR champion outperforming the heterogeneous ensembles in classification accuracy for Mandera and Turkana respectively. This observation makes for the caution in the use of model ensembles as in some tasks, loss of performance can be realized. This loss for performance was observed in the estimation of software efforts in Elish (2013) and in Kocaguneli, Kultur & Bener (2009). The study in Elish (2013) had a voting ensemble model derived from five techniques outperforming single models of the techniques in only three out of five datasets. Kocaguneli, Kultur & Bener (2009) saw heterogeneous ensembles realize accuracies that were far from outperforming single learners.

The building of models that are not specific to each spatial unit of analysis provides an opportunity to scale this approach to the prediction of vegetation conditions across multiple spatial extents. Building spatially down scalable models together with the illustrated generalizability with time as modelled using test data from the future makes this approach highly generalizable both spatially and temporally. This is as opposed to models fine-tuned for each spatial unit separately as were for example reported in Nay, Burchfield & Gilligan (2018).

The superiority in performance between ANN and SVR and indeed between any two machine learning techniques is one on which the jury is still out. In this study, the ANN technique is shown to generally outperform SVR in predictive performance in the ratio of 43% to 5% and a tie in 52% of the cases respectively. This superiority of the ANN to SVR is also observed in the prediction of pipe burst rates in Shirzad, Tabesh & Farmani (2014). Other studies like that in Mokhtarzad et al. (2017) document the SVR to be superior to the ANN. Depending on the application of the techniques, it remains non-clear to pick out one between ANN and SVR as superior to the other. What is clear in this study, however, is that the ANN champion model outperforms the SVR champion in the unseen data even though they remain competitive in model training.

The ANN technique is particularly documented to be more susceptible to overfitting in cases of large networks. This is however not the case in this study as it produces mixed results. The prediction of drought severity documents overfitting as more pronounced in SVR as compared to ANN at 9% to 6%. This result should be interpreted also in the context that overfitting remains confined to only models with $R^2 \leq 0.5$ in SVR as compared to the ANN technique that has two overfit models with $R^2 \geq 0.7$. A contrasting set of results was however reported from the prediction of drought effects on nutrition conditions where overfitting was an exclusive occurrence in the ANN technique with 27% of ANN models reporting overfitting as compared to the non-occurrence of overfitting in the SVR technique. In both instances, we judge overfitting to have been more pronounced amongst the less predictive models as compared to the normal occurrence of overfitting amongst models that are generally judged to be highly predictive. The reduced occurrence of overfitting could be attributed to the use of a sample size that was adequate in the model training process, thus avoiding the high dimension, low sample size learning scenarios as documented in Liu et al. (2017).

Finally, the prediction of nutrition status 1 month ahead for children under 5 years old gives a promising performance. With 272 models of ANN and a similar number of SVR models realised with $R^2 \geq 0.7$ and at the same time not overfit, there is potential to develop enough models for ensembling using the approach of this study. As the case for drought severity, the stacked approach is shown to be superior to the other model ensembling approaches in the prediction of drought effects. In regression, the 8% and 11% lead for homogeneous stacked ANN and SVR respectively offer enough motivation for the use of this approach in realizing highly predictive models. It is, however, the case that for classification, stacking remains competitive to other approaches, perhaps due to the large bands used to realise the classes.

Chapter 5: CONCLUSION

5.1 Summary of the research

The prediction of drought is one of the fundamental improvements to drought early warning systems (DEWS). The demand for drought early warning systems that incorporate prediction is ever increasing. This demand for predictive systems is in congruence with the need not just for such systems but in a context where they offer highly accurate and stable predictions of future drought conditions. This study produced base models of both ANN and SVR that were then investigated for performance in term of champion models, homogeneous ensembles and heterogeneous ensembles. The drought models also aimed to solve the tendency to build predictive models only using only one index or a group of indexes not covering the entire spectrum of drought that is made up of meteorological, hydrological, agricultural and socio-economic droughts. The achievements of this study were majorly documented in two chapters beginning with methodology and ending with the chapter on results and discussion.

Chapter one presented the introduction to the study. The section on introduction covered the background to the study that outlined the economic, social, environmental, policy and technical and research views to the problem of drought. The current state of the art in drought prediction and in drought monitoring and the future expectations from drought early warning systems (DEWS) were documented. The introduction also formulated the problem statement of the study and anchored it in the study objectives and research questions. The objective of the study revolved around the investigation of all variables used in the study of drought and the role of model ensembling in the improvement of both model accuracy and model stability.

The chapter on literature review anchored the research on theory, identified datasets used in the study and was the basis of evidence that supports the findings of this study. The chapter reviewed the definitions of drought establishing the non-agreement on a single definition and view of the phenomenon. The types of drought we reviewed and methods used in the prediction of droughts were reviewed. The chapter documented the emerging trends in drought management especially on ensemble modelling and the increasing need for integration of socio-economic data for the ground-truthing of the effects of drought. The chapter also reviewed model ensembling and ensembling methods as an emerging trend in the development of predictive models even in fields away from drought monitoring. The chapter on literature review finally outlined the conceptual framework concluding with the identification of variables to be used in the study of the prediction of both drought severity and drought effects.

The third chapter dealt with the question of research design and methodology. The correlational research design was exhaustively documented and justification as to why other research designs were not suitable choices for the study provided. Data collection and pre-processing for both remote sensing and socio-economic data was covered in this section including the need for automated process in the download, sub-setting, smoothing and aggregating of the remote sensing datasets. The machine learning techniques for regression were reviewed with the case study techniques of ANN and SVR detailed in-terms of their methodology and process of use. Model ensembling approaches were reviewed including the ensembling types – both homogeneous and heterogeneous.

Chapter four presented the results from the study organised by the study objectives. The next sections of this chapter review the significance of the study, the key achievements under each objective, generalization, limitations of the study and recommendations for the furtherance of the work.

5.2 Achievements

The study had three objectives each with two research questions except for the second objective that had three research questions. Answering the questions would be interpreted as meeting the objectives of the study. This sub-section presents a summary of the achievements organized by the objectives and questions.

5.2.1 Objective One: *Determine the different biophysical and socio-economic variables that are used in the monitoring/ prediction of drought and investigate their relationship with drought.*

In this objective, we undertook review literature to assemble the set of indicators across all the drought types of meteorological, hydrological, agricultural and socio-economic that have been variously used for drought monitoring and drought prediction. The variables covered all the listed types of drought. Further to this, the relationships amongst these variables were investigated as well as their relationship to drought severity. The achievements under the two research questions are:

5.2.1.1: RQ1: *What are the different biophysical and socio-economic variables that are used in the monitoring/ prediction of drought?*

A total of 16 variables used in drought monitoring as applied to drought severity were identified. Two variables- MUAC and TOT were identified for the quantification of drought impacts on the nutrition of children below 5 years.

5.2.1.2: RQ2: *How do the variables identified for drought monitoring relate with drought?*

The variables identified from RQ1 were investigated for relationships amongst themselves and with drought severity (VCI3M). The first part of the investigation was geared towards the selection between TAMSAT and CHIRPS variables to retain the most predictive of the two datasets.

The study successfully showed that both sets of TAMSAT and CHIRPS datasets were non-normally distributed by both visual inspection and Shapiro-Wilk test. The two datasets remained competitive in the prediction of drought severity based on multiple metrics for variable selection. The multiple set of metrics used included Spearman's correlation coefficient, Stepwise regression, Akaike information criterion and relative importance criterion that ended up with TAMSAT having the best top two ranked variables over those of TAMSAT. Similar results were recorded by modelling using both GAM and SVR techniques. On the results of the above selection criteria, the study settled on TAMSAT as the source of precipitation data.

On correlation with drought severity, the vegetation condition variables were shown to generally have a higher correlation with drought severity with lagged values of VCI3M, VCI1M and VCIdekad having correlation coefficients above 0.8 ($r>0.8$). SPEI1M and SPEI3M had the lowest correlations with drought severity ($r<0.3$). The temperature condition index aggregated over the last one month (TCI1M) had a strong negative correlation to drought severity. The variables, therefore, exhibited different relationships with drought severity and were capable of offering value in the prediction of drought severity.

5.2.2 Objective Two: To "*build and evaluate the performance of multiple models for drought prediction using Artificial Neural Networks (ANN) and Support Vector Regression (SVR) as the case study Machine Learning methods",* the study built multiple ANN and SVR models from which model ensembles were built. The achievements of each of the research questions are as outlined in 5.2.2.1 and 5.2.2.2 respectively.

5.2.2.1: ***RQ3:*** *What are the multiple models of both Artificial Neural Networks (ANN) and Support Vector Regression (SVR) that can be built for the prediction of both drought severity and drought effects?*

The study, by experimentation, defined the configuration of 2-5-3-1 for the prediction of drought severity. A total of 143 models were built using both ANN and SVR from the sized down modelling space through an automated process using a set of R scripts. The ANN and SVR techniques were presented with similar but randomly partitioned 10 sub-sets of the training data. A total of 111 models and 272 models each from the ANN and SVR techniques that had $R^2 > 0.7$ in the validation dataset and were not overfit were then chosen for model ensembling in the prediction of drought severity and drought effects respectively.

5.2.2.2: ***RQ4:*** *What is the performance of the ANN models as compared to SVR models in the prediction of drought severity?*

The study exhaustively investigated the performance of the ANN and SVR models in the prediction of drought severity using multiple performance metrics. The prediction tasks were each investigated as both a regression problem and subsequently in the operational application as a classification problem.

The problem of loss of performance for single champion models is elaborated by the fact that in model validation both the ANN and SVR had the same performance (R^2=0.83) as regressors while in the testing dataset (2016-2017), the SVR champion lost performance to post an R^2 of 0.78. This potential loss of performance by single champion models in the prediction of drought severity justifies model ensembling.

The study successfully showed that champion models for the prediction of drought severity from the ANN and SVR techniques are predictive and therefore the techniques remain appropriate and are capable of handling the non-linear nature of the remote sensing. While the ANN technique is susceptible to overfitting, it produced stable models capable of offering good predictive power in out of sample datasets.

5.2.2.3: ***RQ5:*** *What is the performance of the ANN models as compared to SVR models in the prediction of drought effects?*

Like was the case for drought severity, from the total of 976 models, we established the superiority of the SVR technique over the ANN technique. More SVR models had R^2 of 0.7 and above as compared to the ANN models. Even reduced to the basic threshold of above chance models with $R^2 \geq 0.5$, SVR marginally beat ANN by 488 to 487 models. An interesting result was the fact that despite the better performance of the SVR technique in generating many models considered predictive, the ANN champion model still outperformed the SVR champion model in model validation and model testing.

5.2.3 Objective Three: To *"build and evaluate the performance of homogeneous and heterogeneous ensemble models of both ANN and SVR in the prediction of drought severity and drought effects.",* the study built homogeneous and heterogeneous model ensembles for the prediction of both drought severity and drought effects and investigated their performance as outline in 5.2.3.1 and 5.2.3.2 respectively.

5.2.3.1: ***RQ6:*** *What is the performance of the Artificial Neural Networks (ANN) and Support Vector Regression (SVR) homogeneous ensemble models in the prediction of both drought severity and drought effects?*

The ANN and SVR techniques each had 111 models chosen for model ensembling. Each technique had its 111 models ensembled separately using three approaches to model ensembling.

The study showed the superiority of stacking as a homogeneous model ensembling approach in the prediction of both drought severity and drought effects. In regression, model stacking had performance gains from R^2=0.82 to R^2 =0.91 for the ANN technique and from R^2=0.78 to R^2 =0.88 for the SVR technique. As classifiers, the AUROC for each case improved from 89.37% to 91.26% for the ANN and from 88.54 to 91.20 for the SVR techniques respectively. Similar results were posted in the prediction of drought effects with the homogeneous ANN ensembles improving in performance from 0.74 to 0.82 in R^2 using the stacking approach while SVR homogeneous improved from 0.71 to 0.77.

5.2.3.2: ***RQ7:*** *What is the performance of the ANN and SVR heterogeneous ensemble models in the prediction of drought severity and drought effects?*

In the prediction of drought severity, stacking of heterogeneous models improved the performance of the champion models from R^2 of 0.82 for ANN and 0.78 for SVR to an R^2 of 0.94. This as compared to R^2 of 0.82 for both simple and weighted averaging.

In the prediction of drought effects, the stacked approach to building ensembles still out-performed the other approaches. The heterogeneous ensembles out-performed the homogeneous ensembles except for stacking that returns the same R^2 of 0.82. In summary, the study achieved to show that:

- Any model ensembling outperforms the champion models in both regression and classification and also realizes relatively more stability in the testing data.
- Stacking guarantees better performance as compared to non-weighted simple averaging and rank weighted averaging.

- In the prediction of both drought severity and drought effects, heterogeneous ensembles out-perform both homogeneous ensembles and single champion models in both regression and classification.
- The simple and weighted averaging approaches at times do not guarantee an improvement in performance as evidenced by the case recorded by the study.

5.3 Contributions

The contributions of the study have been grouped as theoretical, practical or methodological.

5.3.1 Theoretical Contribution

The study investigated relationships between vegetation datasets processed using different approaches. The contribution of this investigation is a guide to the choice of pre-processing steps. With different datasets from the same base MODIS datasets, we have shown pre-processing steps do make a difference but the statistical test for agreements show very high correlations amongst the datasets. In this aspect, the study bridged the gap between theory and practice by actual comparison of competing datasets prior to a model building process rather than the approach of use of non-objectively chosen data sets.

The study reviewed existing literature and documented the tendency of drought monitoring studies to focus on one type of drought from the set of meteorological, hydrological, agricultural and socio-economic droughts. We make a theoretical contribution to the use of multiple indicators covering the whole set of drought types.

The inclusion of socio-economic data in the predictive study is useful especially in ground-truthing the effects of drought on elements at risk. The theory on the development of predictive models, for both drought severity and drought effects, is a key formulation of this study. While the collection of socio-economic data is non-trivial, their use in drought management amounts to ground validation of drought impacts.

The study delivered an evaluation premise that is appropriate in the determination of the best methods for the development of model ensembles. The question on the performance of model ensembles based on the competing methods of linear average, rank weighted average and model stacking proved the theoretical assumption of the superiority of model stacking.

5.3.2 Methodological Contribution

The methodological contributions are in multiple aspects of the processes that delivered this study. First, as opposed to most methods, we developed model ensembles and compared them to the common practice of building single champion models. The methodology also investigated, empirically, the performance of both homogeneous and heterogeneous model ensembles using multiple machine learning techniques of artificial neural networks (ANN) and support vector regression (SVR). Multiple approaches were used in the building of the model ensembles including simple and weighted averaging and model stacking using ANN perceptron learnt weights. A settlement of the superiority of model stacking in the building of model ensembles is a contribution to both theory and methodology. This is augmented by the use of the statistical General Additive Model (GAM) approach in the reduction of model space complexity using objective multiple model evaluation parameters. The study empirically established the utility of assumptions and the use of GAM technique in model space reduction together with model ensemble membership selection in realising fewer but more predictive ensemble membership.

5.3.3 Technical Contributions

The study delivers a set of R scripts for both the download and pre-processing of the non-vegetation remote sensing data and also for the creating of champion models, homogeneous and heterogeneous model ensembles. Despite some of the steps in the development of the ensembles requiring handcrafting, the automation of the processes ensures the studies can be replicated. The model building process had the script modified to loop through all models in the model space, bootstrap aggregated the training data, build all the ANN and SVR models on the different training data sets and logged the results of model training, validation and evaluation of performance metrics. An additional practical contribution of the research is the publication of two research papers as an output of the research.

5.4 Generalization of results

Two questions on the models produced are if they can be applied in areas not specific to the area of study and if the models will perform well in the future. The ability of the location-specific models to be applicable in other areas is referred to as transferability while the ability of the models to perform well in previously unseen data is referred to as generalization.

Apart from the socio-economic data used in the prediction of drought effects, the other datasets can be sourced from open source remote sensing data repositories. Appropriate replacement of the TAMSAT datasets with any global dataset will ensure the results can be replicated globally using the methodology and datasets chosen. In fact, for the African continent, the only effort would be to avail the appropriate demarcation of a study area preferably as a shapefile. The investigation of the prediction of drought effects will, however, demand the search for data on terms of trade from markets monitoring and such other market performance data and also data on Mid-Upper Arm Circumference even from hospital admission in some instances.

The question on generalization was well handled by the study methodology on the building and validation of models. The out of sample dataset for 24 months across the counties in the study areas showed good generalizability at R^2 of 0.83 for the ANN, 0.78 for the SVR and up to 0.94 for the stacked heterogeneous model ensembles. The delivered models and application of the techniques resulted in a methodology that is generalizable both spatially, hence by location, and temporally into the future.

5.5 Limitations

There is opportunity to do multiple predictions with different lead times so as to offer longer time frames for drought preparedness and better intervention planning. An example would be the possibility of offering 3-months and possibly 6-months predictive models. These will, however, require a longer temporal coverage of the historical data since lagging the variables reduces the data available for the training of the models.

The study chose two machine learning techniques. The two techniques of Artificial Neural Networks (ANN) and Support Vector Regression (SVR) were chosen on the basis of their appropriateness for regression and their ability to model non-linear decision surfaces. The techniques have outputs that are non-linear with respect to the model inputs. The study could benefit from more algorithms that have these characteristics and that can then make for more than two techniques in the investigation of heterogeneous model ensembling.

The investigation of drought effects was done using two variables- MUAC and Terms of Trade (ToT). The use of ToT was to reduce the model space complexity by reducing the number of possible models to be computed and hence ensembled. The set of models for prediction of drought effects was also built as a combination of the two socio-economic variables and the models deemed to have been predictive of drought severity. Though this set up is logical since it links drought effects to drought severity, even the other models left out could be good performers in the prediction of drought effects. Computational power permitting, the model space could be brute-forced as a whole in the realization of all possible predictive models for drought effects.

5.6 Opportunities for Operational drought monitoring

The opportunity for operational drought monitoring from this research can be viewed in two aspects. First, the predictive model can be used with some lead time for drought response planning. The predictions being for both drought severity and drought effects make for the incorporation of ground truths in both the monitoring and prediction of drought. Second, is the fact that the study used multiple indicators and multiple indexes in the prediction of future drought conditions using multiple models whose performance was investigated based on multiple ensemble methods.

The approach to prediction using model ensembles overcomes the tendency of champion models to lose predictive power as was illustrated with the case of the champion SVR model that lost performance from an R^2 of 0.83 to 0.78 between validation and test datasets. Even in the case of the SVR model, the second-best model in training and validation was able to maintain performance at an R^2 of 0.83 in both validation and testing datasets. Ensembling overcomes this limitation of the choice of a single best performer model and offers better predictive power. Such highly predictive and stable models realized from model ensembling build confidence in model outputs especially in real-life applications that involve resource allocations.

5.7 Future Work

One critical variable for hydrological drought that was not included in the study but provides for a future opportunity is streamflow index. The streamflow index was not used as a result of the non-existence of large water bodies in the study area except for Turkana county.

The section on the limitations of the study documented an opportunity in the development of multiple periods for predictions through longer lags of the variables. Sample periods would, for example, be 6 months and 12 months lead times for better drought preparedness.

An additional opportunity for enhancement would be to work with the players in the food security sectors to document most of the indicators that are used in food security assessment and that are deemed important for drought monitoring. Augmenting the identified socio-economic data with those from food security would offer longer-term time series data.

Finally, models predicting multiple aspects of the effects of droughts like say prices together with malnutrition and possibly others like crop yields would be a natural progression of this study.

References

Adams, P. D., Horridge, M., Madden, J., & Wittwer, G. (2002). Drought Regions and the Australian Economy between 2001-2 and 2004-5.

Adede, C., Oboko, R., Wagacha, P. W., & Atzberger, C. (2019a). A Mixed Model Approach to Vegetation Condition Prediction Using Artificial Neural Networks (ANN): Case of Kenya's Operational Drought Monitoring. *Remote Sensing*, *11*(9), 1099.

Adede, C., Oboko, R., Wagacha, P. W., & Atzberger, C. (2019b). Model Ensembles of Artificial Neural Networks and Support Vector Regression for Improved Accuracy in the Prediction of Vegetation Conditions and Droughts in Four Northern Kenya Counties. *ISPRS International Journal of Geo-Information*, *8*(12), 562.

Agana, N. A., & Homaifar, A. (2017). A hybrid deep belief network for long-term drought prediction. In *Proceedings of the Workshop on Mining Big Data in Climate and Environment (MBDCE 2017), 17th SIAM International Conference on Data Mining (SDM 2017), Houston, TX, USA* (pp. 27-29).

AghaKouchak, A., & Nakhjiri, N. (2012). A near real-time satellite-based global drought climate data record. *Environmental Research Letters*, *7*(4), 044037.

AghaKouchak, A., Farahmand, A., Melton, F. S., Teixeira, J., Anderson, M. C., Wardlow, B. D., & Hain, C. R. (2015). Remote sensing of drought: Progress, challenges and opportunities. *Reviews of Geophysics*, *53*(2), 452-480.

Albarakat, R., & Lakshmi, V. (2019). Comparison of Normalized Difference Vegetation Index Derived from Landsat, MODIS, and AVHRR for the Mesopotamian Marshes Between 2002 and 2018. *Remote Sensing*, *11*(10), 1245.

Ali, Z., Hussain, I., Faisal, M., Nazir, H. M., Hussain, T., Shad, M. Y., ... & Hussain Gani, S. (2017). Forecasting drought using multilayer perceptron artificial neural network model. *Advances in Meteorology*, *2017*.

Anderson, R. (2007). *The Credit Scoring Toolkit*: Oxford University Press.

Anuradha, J. (2015). A brief introduction on Big Data 5Vs characteristics and Hadoop technology. *Procedia computer science*, *48*, 319-324.

Atzberger, C., Carter, M., Fava, F., Jensen, N., Meroni, M., Mude, A., ... & Vrieling, A. (2016). Does the Design Matter. *Comparing Satellite-Based Indices for Insuring Pastoralists in Kenya: Technical Report Prepared for the BASIS Assets and Market Access CRSP*, 2017-05.

Azevedo, A. and Santos, M.F. (2008). KDD, SEMMA and CRISP-DM: A parallel overview. *Proceedings of the AIDIS European Conference on Data Mining*, pp. 182-185.

Bachmair, S., Kohn, I., & Stahl, K. (2015). Exploring the link between drought indicators and impacts. *Nat. Hazards Earth Syst. Sci*, *15*(6), 1381-1397.

Bachmair, S., Stahl, K., Collins, K., Hannaford, J., Acreman, M., Svoboda, M., ... & Crossman, N. D. (2016). Drought indicators revisited: the need for a wider consideration of environment and society. *Wiley Interdisciplinary Reviews: Water*.

Basara, J. B., Christian, J. I., Wakefield, R. A., Otkin, J. A., Hunt, E. H., & Brown, D. P. (2019). The evolution, propagation, and spread of flash drought in the Central United States during 2012. *Environmental Research Letters*, *14*(8), 084025.

Batini, C., Blaschke, T., Lang, S., Albrecht, F., Abdulmutalib, H. M., Barsi, Á., ... & Kugler, Z. (2017). Data Quality in Remote Sensing. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *42*.

Beesley, J. (2011). The Hunger Safety Nets Programme in Kenya: A Social Protection Case Study. *Oxfam GB*.

Beguería, S., Vicente-Serrano, S. M., Reig, F., & Latorre, B. (2014). Standardized precipitation evapotranspiration index (SPEI) revisited: parameter fitting, evapotranspiration models, tools, datasets and drought monitoring. *International Journal of Climatology*, *34*(10), 3001-3023.

Belayneh, A., & Adamowski, J. (2012). Standard precipitation index drought forecasting using neural networks, wavelet neural networks, and support vector regression. *Applied Computational Intelligence and Soft Computing*, *2012*.

Belayneh, A., Adamowski, J., Khalil, B., & Quilty, J. (2016). Coupling machine learning methods with wavelet transforms and the bootstrap and boosting ensemble approaches for drought prediction. *Atmospheric research*, *172*, 37-47.

Below, R., Grover-Kopec, E., & Dilley, M. (2007). Documenting drought-related disasters a global reassessment. *The Journal of Environment & Development*,*16*(3), 328-344.

Berhan, G., Hill, S., Tadesse, T. &Atnafu, S. (2011). Using satellite images for drought monitoring: A knowledge discovery approach. *Journal of Strategic Innovation and Sustainability*, *7*(1), 135-153.

Berwick, R. (2003). An idiot’s guide to support vector machines. *Advanced Computer Vision. University of Central Florida*.

Bilbao, I., & Bilbao, J. (2017, December). Overfitting problem and the over-training in the era of data: Particularly for Artificial Neural Networks. In *2017 Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)* (pp. 173-177). IEEE.

Bishop, C. M. (2006). *Pattern recognition and machine learning* (Vol. 1, p. 740). New York: springer.

Bordi, I., Fraedrich, K., Petitta, M., & Sutera, A. (2005, September). Methods for predicting drought occurrences. In *Proceedings of the 6th International Conference of the European Water Resources Association, Menton, France* (pp. 7-10).

Brockwell, P. J., & Davis, R. A. (2006). *Introduction to time series and forecasting*. Springer Science & Business Media.

Brown, J., Howard, D., Wylie, B., Frieze, A., Ji, L., & Gacke, C. (2015). Application-ready expedited MODIS data for operational land surface monitoring of vegetation condition. *Remote Sensing*, *7*(12), 16226-16240.

Bulmer, M. G. 1979. *Principles of Statistics*. Courier Corporation.

Bunting, P. (2017). Pre-processing of remotely sensed imagery. In *The Roles of Remote Sensing in Nature Conservation* (pp. 39-63). Springer, Cham.

Chand, K., & Biradar, N. (2017). Socio-Economic Impacts of Drought in India. *Drought Mitigation and Management*, 245.

Chantarat, S., Mude, A. G., Barrett, C. B., & Carter, M. R. (2013). Designing index-based livestock insurance for managing asset risk in northern Kenya. *Journal of Risk and Insurance, 80*(1), 205-237.

Chao, W. L., (2011), 'Machine Learning Tutorial. [Online]. Available at: http://disp.ee.ntu.edu.tw/~pujols/Machine%20Learning%20Tutorial.pdf (Accessed 12 November 2016)

Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. *SPSS inc*, *9*, 13.

Cody, B. A., Folger, P., & Brougher, C. M. (2009). California drought: Hydrological and regulatory water supply issues.

Constitution, K. (2010). Government printer. *Kenya: Nairobi*. [Online]. Available at: https://www.refworld.org/docid/4c8508822.html [Accessed 25 May 2019]

Creswell, J. W. (2013). *Research design: Qualitative, quantitative, and mixed methods approaches*.London:Sage.

Creswell, J. W., & Creswell, J. D. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage publications.

Cristianini, N., & Shawe-Taylor, J. (2000). *An introduction to support vector machines and other kernel-based learning methods*. Cambridge university press.

Curtis, E. A., Comiskey, C., & Dempsey, O. (2016). Importance and use of correlational research. *Nurse researcher*, *23*(6).

Da Silva, J., & Cernat, L. (2012). Coping with loss: the impact of natural disasters on developing countries' trade flows. *EC Chief Economist Note*, (1).

Dalla Mura, M., Prasad, S., Pacifici, F., Gamba, P., Chanussot, J., & Benediktsson, J. A. (2015). Challenges and opportunities of multimodality and data fusion in remote sensing. *Proceedings of the IEEE*, *103*(9), 1585-1601.

De Onis, M., Yip, R., & Mei, Z. (1997). The development of MUAC-for-age reference data recommended by a WHO Expert Committee. *Bulletin of the World Health Organization*, *75*(1), 11.

Demuth, S., & Stahl, K. (2001). *Assessment of the Regional. Impact of Droughts in Europe (ARIDE)-Final Report*. University of Freiburg/Institute of Hydrology.

Di Falco, S., & Veronesi, M. (2013). How can African agriculture adapt to climate change? A counterfactual analysis from Ethiopia. *Land Economics, 89*(4), 743-766.

Didan, K. (2015a). *MOD13Q1 MODIS/Terra Vegetation Indices 16-Day L3 Global 250m SIN Grid V006* [Data set]. NASA EOSDIS LP DAAC. doi: 10.5067/MODIS/MOD13Q1.006

Didan, K. (2015b). *MYD13Q1 MODIS/Aqua Vegetation Indices 16-Day L3 Global 250m SIN Grid V006* [Data set]. NASA EOSDIS LP DAAC. doi: 10.5067/MODIS/MYD13Q1.006

Ding, Y., Hayes, M. J., & Widhalm, M. (2011). Measuring economic impacts of drought: a review and discussion. *Disaster Prevention and Management: An International Journal*.

Dinku, T., Ceccato, P., Grover-Kopec, E., Lemma, M., Connor, S. J., & Ropelewski, C. F. (2007). Validation of satellite rainfall products over East Africa's complex topography. *International Journal of Remote Sensing*, *28*(7), 1503-1526.

Drucker, H., & Cortes, C. (1996). Boosting decision trees. In *Advances in neural information processing systems* (pp. 479-485).

Du, L., Tian, Q., Yu, T., Meng, Q., Jancso, T., Udvardy, P., & Huang, Y. (2013). A comprehensive drought monitoring method integrating MODIS and TRMM data. *International Journal of Applied Earth Observation and Geoinformation*, *23*, 245-253.

Džeroski, S., & Ženko, B. (2004). Is combining classifiers with stacking better than selecting the best one?. *Machine learning*, *54*(3), 255-273.

Elish, M. O. (2013, April). Assessment of voting ensemble for estimating software development effort. In *2013 IEEE Symposium on Computational Intelligence and Data Mining (CIDM)* (pp. 316-321). IEEE.

EMDAT, C. R. E. D. (2012). The international disaster database. *Center for Research on the Epidemiology of Disasters.* [Online]. Available at*:* https://www. emdat.be (Accessed 23 March 2020].

Enenkel, M., See, L., Bonifacio, R., Boken, V., Chaney, N., Vinck, P., ...& Anderson, M. (2015). Drought and food security–Improving decision-support via new technologies and innovative collaboration.*Global Food Security*, *4*, 51-55.

Enenkel, M., Steiner, C., Mistelbauer, T., Dorigo, W., Wagner, W., See, L., ... & Rogenhofer, E. (2016). A combined satellite-derived drought indicator to support humanitarian aid organizations. *Remote Sensing*, *8*(4), 340.

Eslamian, S., Ostad-Ali-Askari, K., Singh, V. P., Dalezios, N. R., Ghane, M., Yihdego, Y., & Matouq, M. (2017). A review of drought indices. *Int J Constr Res Civ Eng (IJRCRE)*, *3*(4), 48-66.

FAO (2019). *Proactive approaches to drought preparedness – Where are we now and where do we go from here.* [Online]. Available at: http://www.fao.org/3/ca5794en/ca5794en.pdf (Accessed 10 March 2020)

Fayyad, U. M., Piatetsky-Shapiro, G. & Smyth, P. (1996), From data mining to knowledge discovery: An overview, in 'Advances in Knowledge Discovery and Data Mining', AAAI Press and the MIT Press, chapter 1, pp. 1-34.

Few, S. (2009). Introduction to geographical data visualization. *Visual Business Intelligence Newsletter*, 1-11 [Online]. Available at: http://www.perceptualedge.com/articles/visual_business_intelligence/geographical_data_visualization.pdf (Accessed 7 January 2015)

FEWSNET (2009). Terms of Trade and Food Security Analysis: FEWS NET Markets Guidance, No 5. [Online] Available at: https://fews.net/sites/default/files/MT%20Guidance_Terms%20of%20Trade_No%205_En.pdf (Accessed 13 April 2020)

FEWSNET Data Portal (2016). USGS FEWS NET Data Portal. [Online] Available at: http://earlywarning.usgs.gov/fews (Accessed 11 October 2016)

Fortmann-Roe, S. (2012). Understanding the Bias-Variance Tradeoff. [Online] Available at: http://scott.fortmann-roe.com/docs/BiasVariance.html#fn:1 (Accessed 30 April 2016)

Funk, C. C., Peterson, P. J., Landsfeld, M. F., Pedreros, D. H., Verdin, J. P., Rowland, J. D., & Pedreros, P. (2014). A quasi-global precipitation time series for drought monitoring. US Geological Survey Data Series, 832(4).

Funk, C., Peterson, P., Landsfeld, M., Pedreros, D., Verdin, J., Shukla, S., & Michaelsen, J. (2015). The climate hazards infrared precipitation with stations—a new environmental record for monitoring extremes. Scientific data, 2.

Fürnkranz, J., Gamberger, D., & Lavrač, N. (2012). *Foundations of rule learning*. Springer Science & Business Media.

Garrido, A. (2014), 'National Capacity Development Training of Trainers (TOTs), paper presented to Workshop on Developing and Implementing Mitigation and Preparedness Water Scarcity and Drought (WS&D) Management Plans, Zaragoza-Madrid, Spain, May 6-9, 2014

Gebremeskel, G., Tang, Q., Sun, S., Huang, Z., Zhang, X., & Liu, X. (2019). Droughts in East Africa: Causes, impacts and resilience. *Earth-science reviews*.

Goebel, M. and Gruenwald, L. (1999). A survey of data mining and knowledge discovery software tools. *Explorations Newsletter*, *1*(1), pp. 20-33*ACM SIGKDD* [Online].Available at: http://www.kdd.org/sites/default/files/issues/1-1-1999-06/survey.pdf (Accessed 18 December 2014)

Government of Kenya (2012). *Kenya Post-Disaster Needs Assessment:* 2008–2011 Drought. [Online]. Available at: http://www.gfdrr.org/sites/gfdrr/files/Kenya_PDNA_Final.pdf(Accessed 09 November 2018)

Government of Kenya (2014). *Ending Drought Emergencies: Common Programme Framework.* [Online]. Available at: https://www.ndma.go.ke/index.php/resource-center/send/43-ending-drought-emergencies/4251-common-programme-framework(Accessed 25 February 2020)

Grasso, V. F., & Singh, A. (2011). Early warning systems: State-of-art analysis and future directions. *Draft report, UNEP*, *1*.

Gu, Y., Hunt, E., Wardlow, B., Basara, J. B., Brown, J. F., &Verdin, J. P. (2008). Evaluation of MODIS NDVI and NDWI for vegetation drought monitoring using Oklahoma Mesonet soil moisture data.*Geophysical Research Letters*, *35*(22).

Gunst, R. F. (2018). *Regression analysis and its application: a data-oriented approach*. Routledge.

Gutiérrez, A. P. A., Engle, N. L., De Nys, E., Molejón, C., & Martins, E. S. (2014). Drought preparedness in Brazil. *Weather and Climate Extremes*, *3*, 95-106.

Hagedorn, R., Doblas-Reyes, F. J., & Palmer, T. N. (2005). The rationale behind the success of multi-model ensembles in seasonal forecasting—I. Basic concept. *Tellus A: Dynamic Meteorology and Oceanography*, *57*(3), 219-233.

Han, J., Kamber, M., & Mining, D. (2006). Data Mining, Southeast Asia Edition: Concepts and Techniques. Burlington: Morgan Kaufmann.

Hand, D. J., Mannila, H., & Smyth, P. (2001). *Principles of data mining*. Cambridge: MIT press.

Hand, D.J., & Till, R.J. (2001). Ample Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Machine Learning* **45**(2), p. 171--186. DOI: 10.1023/A:1010920819831.

Hao, Z., & AghaKouchak, A. (2013). Multivariate standardized drought index: a parametric multi-index model. *Advances in Water Resources*, *57*, 12-18.

Hao, Z., AghaKouchak, A., Nakhjiri, N., & Farahmand, A. (2014). Global integrated drought monitoring and prediction system. *Scientific data*, *1*, 140001.

Hao, Z., Singh, V. P., & Xia, Y. (2018). Seasonal drought prediction: advances, challenges, and future prospects. *Reviews of Geophysics*, *56*(1), 108-141.

Hardle, W. (1990). *Applied nonparametric regression* (Vol. 27). Cambridge: Cambridge university press.

Hastie, T. J. (2017, September). Methods for predicting drought occurrences. In *Statistical Models. In S Routledge: Abingdon, UK* (pp. 249-307).

Hayes, M. J., Svoboda, M. D., Wiihite, D. A., & Vanyarkho, O. V. (1999). Monitoring the 1996 drought using the standardized precipitation index. *Bulletin of the American meteorological society*, *80*(3), 429-438.

Hayes, M., Svoboda, M., Wall, N., & Widhalm, M. (2011). The Lincoln declaration on drought indices: universal meteorological drought index recommended. *Bulletin of the American Meteorological Society*, *92*(4), 485-488.

Hewitt, K. 1997. *Regions at Risk. A Geographical Introduction to Disasters*, Harlow: Longman.

Higgins, J. P., Deeks, J. J., & Altman, D. G. (2009). General principles for dealing with missing data. *Cochrane Handbook*.

Hill, R., & Fuje, H. (2017). *What is the impact of drought on prices? Evidence from Ethiopia*. Mimeo.

Hochrainer, S., Mechler, R., Pflug, G., Linnerooth-Bayer, J., Aaheim, A., & Mestl, H. (2007). Preliminary report on Europe's financial and economic vulnerability to meteorological extremes. *Deliverable to the ADAM project, DA2, 3*.

Howitt, R., Medellin-Azuara, J. & Lund, J. (2014). *Preliminary 2014 Drought Economic Impact Estimates in Central Valley Agriculture*.[Online].Available at: https://watershed.ucdavis.edu/files/biblio/Preliminary_2014_drought_economic_impacts-05192014.pdf (Accessed 10 September 2014)

Howitt, R., Medellín-Azuara, J., MacEwan, D., Lund, J., & Sumner, D. (2015). Economic analysis of the 2015 drought for California agriculture. UC Davis Center for Watershed Sciences.

Huang, G. B. (2003). Learning capability and storage capacity of two-hidden-layer feedforward networks. *IEEE Transactions on Neural Networks*, *14*(2), 274-281.

Huang, Y. F., Ang, J. T., Tiong, Y. J., Mirzaei, M., & Amin, M. Z. M. (2016). Drought forecasting using SPI and EDI under RCP-8.5 climate change scenarios for Langat River Basin, Malaysia. *Procedia Engineering*, *154*, 710-717.

Huete, A., Didan, K., Miura, T., Rodriguez, E.P., Gao, X. & Ferreira, L. G. (2002), Overview of the radiometric and biophysical performance of the MODIS vegetation indices, *Remote Sensing of Environment*, 83(1-2) pp. 195-213 [Online]. Available at:

https://cimss.ssec.wisc.edu/dbs/China2011/Day2/Lectures/MODIS_MOD13_NDVI_referenc.pdf (Accessed 15 October 2014)

Humphries, M. (2013). Missing Data & How to Deal: An overview of missing data. *Population Research Center. University of Texas*. [Online] Available at: https://liberalarts.utexas.edu/prc/_files/cs/Missing-Data.pdf (Accessed 19 September 2016)

ICPAC (2007) Available at: http://www.icpac.net/applications/applications.html (Accessed 14 January 2015)

IGAD (2007).*Disaster Risk Management: Training Manual*.[Online].Available at: http://www.disasterriskreduction.net/fileadmin/user_upload/drought/docs/1-DRM%20Training%20Manual.pdf(Accessed 14 January 2015)

ILRI (2009). *An assessment of the response to the 2008-2009 drought in Kenya: A report to the European Union Delegation to the Republic of Kenya*.[Online]. Available at:*http://eeas.europa.eu/delegations/kenya/documents/more_info/assessment_of_the_response_2008_en.pdf*(Accessed 19 January 2015)

Ivey, J. (2015). How important is a conceptual framework?. *Pediatric nursing*, *41*(3), 145.

Jabareen, Y. (2009). Building a conceptual framework: philosophy, definitions, and procedure. *International Journal of Qualitative Methods*, *8*(4), 49-62.

James, W. P., Mascie-Taylor, G. C., Norgan, N. G., Bistrian, B. R., Shetty, P. S., & Ferro-Luzzi, A. (1994). The value of arm circumference measurements in assessing chronic energy deficiency in Third World adults. *European journal of clinical nutrition*, *48*(12), 883-894.

Jeff, S. (2015) 7 Ways to Handle Missing Data. [Online] Available at: http://www.measuringu.com/blog/handle-missing-data.php (Accessed 13 September 2016)

Jenkins, K. L. (2012). *Modelling the economic and social consequences of drought under future projections of climate change* (Doctoral dissertation, University of Cambridge).

Jensen, N., Stoeffler, Q., Fava, F., Vrieling, A., Atzberger, C., Meroni, M., ... & Carter, M. (2019). Does the design matter? Comparing satellite-based indices for insuring pastoralists against drought. *Ecological economics*, *162*, 59-73.

Ji, L. & Gallo, K. (2006). An agreement coefficient for image comparison.*Photogrammetric Engineering & Remote Sensing*, *72*(7), pp. 823-833 [Online] Available at: http://www.researchgate.net/profile/Lei_Ji3/publication/260138053_The_Agreement_Coefficient_for_Image_Comparison/links/53ebc6140cf24f241f15005b.pdf (Accessed 22 December 214)

Jin, L. (2011). A data-driven test to compare two or multiple time series. *Computational Statistics & Data Analysis*, *55*(6), 2183-2196.

Kala, C. P. (2017). Environmental and socioeconomic impacts of drought in India: lessons for drought management. *Sci Educ*, *5*(2), 43-48.

Kala, C. P., & Silori, C. S. (Eds.). (2013). *Biodiversity, communities and climate change*. The Energy and Resources Institute (TERI).

Keshavarz, M., Karami, E., & Vanclay, F. (2013). The social experience of drought in rural Iran. *Land Use Policy*, *30*(1), 120-129.

Khadr, M. (2016). Forecasting of meteorological drought using Hidden Markov Model (case study: The upper Blue Nile river basin, Ethiopia). *Ain Shams Engineering Journal*, *7*(1), 47-56.

Khaleghi, B., Khamis, A., Karray, F. O., & Razavi, S. N. (2013). Multisensor data fusion: A review of the state-of-the-art. *Information fusion*, *14*(1), 28-44.

Khamala, E. (2017). *Review of the available remote sensing tools, products, methodologies and data to improve crop production forecasts*. Food and Agriculture Organization of the United Nations (FAO).

Klisch, A., & Atzberger, C. (2016). Operational Drought Monitoring in Kenya Using MODIS NDVI Time Series. *Remote Sensing*, *8*(4), 267.

Klisch, A., Atzberger, C., & Luminari, L. (2015). Satellite-Based Drought Monitoring in Kenya In an Operational Setting. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*.

KNBS (2018). Basic report on well-being in Kenya. [Online] Available at: http://statistics.knbs.or.ke/nada/index.php/catalog/88 (Accessed 20 January 2019)

KNBS (2019). 2019 Kenya Population and Housing Census Volume I: *Population by County and Sub-county*. [Online]. Available at: https://www.knbs.or.ke/?page_id=3142 (Accessed 16 April 2020)

Kocaguneli, E., Kultur, Y., & Bener, A. (2009, November). Combining multiple learners induced on multiple datasets for software effort prediction. In *20th international symposium on software reliability engineering (ISSRE)*.

Kock, N., Lynn, G. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the Association for Information Systems*, *13*(7).

Kogan, F.N., (1990). Remote sensing of weather impacts on vegetation in non-homogeneous areas, *International Journal of Remote Sensing*, 11(8), pp. 1405-1419.

Kohavi, R. (2001, February). Data mining and visualization. In *Sixth Annual Symposium on Frontiers of Engineering* (pp. 30-40). National Academy Press, DC.

Kubiak-Wójcicka, K., & Bąk, B. (2018). Monitoring of meteorological and hydrological droughts in the Vistula basin (Poland). *Environmental monitoring and assessment*, *190*(11), 691.

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of statistical software*, *28*(5), 1-26.

LP DAAC (2004). Land Processes Distributed Active Archive Center. *USGS Earth Resources Observation and Science (EROS) Center*. [Online] Available at: https://lpdaac.usgs.gov/documents/308/LDOPE_Users_Manual.pdf (Accessed November 2015).

LAADS DAAC (2016). Level-1 and Atmosphere Archive & Distribution System (LAADS) Distributed Active Archive Center (DAAC).[Online] Available at: https://ladsweb.nascom.nasa.gov/ (Accessed October 2016).

Langley, P. & Sage, S. (1994, July). Induction of selective Bayesian classifiers. In *Proceedings of the Tenth international conference on Uncertainty in artificial intelligence.* pp. 399-406. Burlington: Morgan Kaufmann.

Le, M. H., Perez, G. C., Solomatine, D., & Nguyen, L. B. (2016). Meteorological drought forecasting based on climate signals using artificial neural network–a case study in Khanhhoa Province Vietnam. *Procedia Engineering, 154,* 1169-1175.

Lefsky, M. A., & Cohen, W. B. (2003). Selection of remotely sensed data. In Remote *sensing of forest environments* (pp. 13-46). Springer, Boston, MA.

Liu, B., Wei, Y., Zhang, Y., & Yang, Q. (2017, August). Deep Neural Networks for High Dimension, Low Sample Size Data. In *IJCAI* (pp. 2287-2293).

Liu, F., Qin, Q., & Zhan, Z. (2012). A novel dynamic stretching solution to eliminate saturation effect in NDVI and its application in drought monitoring.*Chinese Geographical Science*, *22*(6), 683-694.

Liu, P. (2015). A survey of remote-sensing big data. *frontiers in Environmental Science*, *3*, 45.

Lloyd-Hughes, B. (2014). The impracticality of a universal drought definition. *Theoretical and Applied Climatology*, *117*(3-4), 607-611.

LP DAAC (2014). Land Processes Distributed Active Archive Center. *USGS Earth Resources Observation and Science (EROS) Center*. [Online] Available at: https://lpdaac.usgs.gov/dataset_discovery/modis (Accessed March 8 2016)

Lu, J., Xue, S., Zhang, X., Zhang, S., & Lu, W. (2014). Neural Fuzzy Inference System-Based Weather Prediction Model and Its Precipitation Predicting Experiment. *Atmosphere*, *5*(4), pp, 788-805 [Online] Available at: www.mdpi.com/2073-4433/5/4/788/pdf (Accessed 5 January 2015).

Maca, P., & Pech, P. (2016). Forecasting SPEI and SPI drought indices using the integrated artificial neural networks. *Computational intelligence and neuroscience*, *2016*.

Macdonald, S., & Headlam, N. (2008). *Research Methods Handbook: Introductory guide to research methods for social research*. Centre for Local Economic Strategies.

Madadgar, S., & Moradkhani, H. (2013). A Bayesian framework for probabilistic seasonal drought forecasting. *Journal of Hydrometeorology*, *14*(6), 1685-1705.

Magno, R., De Filippis, T., Di Giuseppe, E., Pasqui, M., Rocchi, L., & Gozzini, B. (2018). Semi-automatic operational service for drought monitoring and forecasting in the Tuscany region. *Geosciences, 8*(2), 49.

Maidment, R. I., Grimes, D., Allan, R. P., Tarnavsky, E., Stringer, M., Hewison, T., ... & Black, E. (2014). The 30 year TAMSAT African rainfall climatology and time series (TARCAT) data set. *Journal of Geophysical Research: Atmospheres, 119*(18), 10-619.

Maidment, R. I., Grimes, D., Black, E., Tarnavsky, E., Young, M., Greatrex, H., ... & Alcántara, E. M. U. (2017). A new, long-term daily satellite-based rainfall dataset for operational monitoring in Africa. *Scientific data*, *4*, 170063.

Mainardi, S. (2011). Cropland use, yields, and droughts: spatial data modeling for Burkina Faso and Niger. *Agricultural Economics, 42*(1), pp. 17-33 [Online]

Available at: https://www.agriskmanagementforum.org/sites/agriskmanagementforum.org/files/Documents/cropland%20use%20burkina%20faso%20and%20niger.pdf (Accessed 8 January 2015)

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A. H. (2011).McKinsey & Company: Big data: The next frontier for innovation, competition, and productivity. Available at: http://www.mckinsey.com/~/media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx (Accessed 6 January 2015).

Mariotti, A., Schubert, S., Mo, K., Peters-Lidard, C., Wood, A., Pulwarty, R., ... & Barrie, D. (2013). Advancing drought understanding, monitoring, and prediction. *Bulletin of the American Meteorological Society*, *94*(12), ES186-ES188.

Mariscal, G., Marbán, Ó. and Fernández, C. (2010). A survey of data mining and knowledge discovery process models and methodologies. *The Knowledge Engineering Review*, *25*(02), pp. 137-166[Online]. Available at: http://www.researchgate.net/publication/220254274_A_survey_of_data_mining_and_knowledge_discovery_process_models_and_methodologies/links/0912f50a8e22278bf8000000.pdf

Martínez-Beltrán, C., Jochum, M. O., Calera, A., & Melia, J. (2009). Multisensor comparison of NDVI for a semi-arid environment in Spain. *International Journal of Remote Sensing*, *30*(5), 1355-1384.

Mas, J. F. & Flores, J. J. (2008). The application of artificial neural networks to the analysis of remotely sensed data. *International Journal of Remote Sensing*, *29*(3), 617-663.

Masinde, M. (2013). Artificial neural networks models for predicting effective drought index: Factoring effects of rainfall variability. *Mitigation and Adaptation Strategies for Global Change*, pp 1-24.[Online]. Available at: http://link.springer.com/article/10.1007%2Fs11027-013-9464-0 (Accessed 12 January 2015)

Massarutto, A., Musolino, D., Pontoni, F., de Carli, A., Senn, L., de Paoli, L., Castro Rego, F., Dias, S., Bifulco, C., Acacio, V. and Andreu, J. (2013). Analysis of historic events in terms of socio-economic and environmental impacts. *DROUGHT-R&SPI Technical Report*, (9).

Mathieson, K., Peacock, E., and Chin, W. W. (2001). "Extending the Technology Acceptance Model: The Influence of Perceived User Resources," *Data Base for Advances in Information Systems* (32:3), pp. 86-112.

McLeod, S. A. (2018, Jan 14) *Correlation*. [Online]. Available at https://www.simplypsychology.org/correlation.html (Accessed 26 June 2018)

McMullan, C., Simic, M., Tosco, A., & Latchman, S. (2016, April). Are Global Economic Losses from Natural Hazards Increasing?. In *EGU General Assembly Conference Abstracts* (Vol. 18, p. 13575).

McVicar, T. R. and Jupp, D. L. (1998). The current and potential operational uses of remote sensing to aid decisions on drought exceptional circumstances in

Australia: a review. *Agricultural systems*, *57*(3), pp. 399-468 [Online] Available at: http://www.researchgate.net/profile/Tim_Mcvicar/publication/221952042_The_current_and_potential_operational_uses_of_remote_sensing_to_aid_decisions_on_drought_exceptional_circumstances_in_Australia_a_review/links/0f317532fcc488a1b1000000.pdf (Accessed 28 August 2014)

Meade, N., & Islam, T. (1998). Technological forecasting—Model selection, model stability, and combining models. *Management science*, *44*(8), 1115-1130.

Mena, J. (1999). Data Mining Your Website. Woburn: Digital Press.

Mendes-Moreira, J., Soares, C., Jorge, A. M., & Sousa, J. F. D. (2012). Ensemble approaches for regression: A survey. *Acm computing surveys (csur)*, *45*(1), 1-40.

Meroni, M., Atzberger, C., Vancutsem, C., Gobron, N., Baret, F., Lacaze, R. and Leo, O. (2013). Evaluation of agreement between space remote sensing SPOT-VEGETATION fAPAR time series. *Geoscience and Remote Sensing, IEEE Transactions on*, 51(4), pp.1951-1962 Available at: http://www.rali.boku.ac.at/fileadmin/data/H03000/H85000/H85700/publications/15_Meroni_et_al_2012_TGRS.pdf (Accessed 20 January 2015)

Meroni, M., Fasbender, D., Rembold, F., Atzberger, C., & Klisch, A. (2019). Near real-time vegetation anomaly detection with MODIS NDVI: Timeliness vs. accuracy and effect of anomaly computation options. *Remote Sensing of Environment*, *221*, 508-521.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Sage.

Mishra, A. K., & Desai, V. R. (2006). Drought forecasting using feed-forward recursive neural network. *ecological modelling*, *198*(1-2), 127-138.

Mishra, A. K., & Singh, V. P. (2011). Drought modeling–A review. *Journal of Hydrology*, *403*(1-2), 157-175.

Mitchell, T. M. (1997). Machine learning. New York: McGraw Hill

Mitchell, T. R. (1985). An evaluation of the validity of correlational research conducted in organizations. *Academy of Management Review*, *10*(2), 192-205.

Mohajan, H. K. (2014). Food and nutrition scenario of Kenya. *American Journal of Food and Nutrition*, *2* (2), 28-38

Mohajan, H. K. (2017). Two criteria for good measurements in research: Validity and reliability. *Annals of Spiru Haret University. Economic Series*, *17*(4), 59-82.

Mokhtarzad, M., Eskandari, F., Vanjani, N. J., & Arabasadi, A. (2017). Drought forecasting by ANN, ANFIS, and SVM and comparison of the models. *Environmental Earth Sciences*, *76*(21), 729.

Monacelli, G., Galluccio, M., & Abbafati, M. (2005). Drought within the context of the Region VI. *WMO, Working group of Hydrology*.

Moody, A., & Johnson, D. M. (2001). Land-surface phenologies from AVHRR using the discrete Fourier transform. *Remote Sensing of Environment*, *75*(3), 305-323.

Morid, S., Smakhtin, V., & Bagherzadeh, K. (2007). Drought forecasting using artificial neural networks and time series of drought indices. *International Journal of Climatology: A Journal of the Royal Meteorological Society*, *27*(15), 2103-2111.

Morton, J., Barton, D., Collinson, C., & Heath, B. (2005). Comparing drought mitigation interventions in the pastoral livestock sector. *Nat. Resour. Institute report, Greenwich, NRI*.
Motha, R., Wilhite, D., & Wood, D. (2011). *Agricultural Drought Indices. Proceedings of an Expert Meeting: 2-4 June, 2010, Murcia, Spain*. M. V. Sivakumar (Ed.). WMO.
Mountrakis, G., Im, J., Ogole, C. (2011). Support vector machines in remote sensing: A review. *ISPRS J. Photogramm. Remote Sens. 2011*, 66, 247–259 Remote Sens. 2011, 66, 247–259.
Moustris, K. P., Nastos, P. T., Larissi, I. K., & Paliatsos, A. G. (2012). Application of multiple linear regression models and artificial neural networks on the surface ozone forecast in the greater Athens area, Greece. *Advances in Meteorology*, *2012*.
Mude, A. G., Chantarat, S., Barrett, C. B., Carter, M. R., Ikegami, M., & McPeak, J. G. (2010). Insuring against drought-related livestock mortality: Piloting index based livestock insurance in northern Kenya. *Available at SSRN 1844745*.
Musial, J. P., Verstraete, M. M., & Gobron, N. (2011). Comparing the effectiveness of recent algorithms to fill and smooth incomplete and noisy time series. *Atmos. Chem. Phys. Discuss*, *11*(14), 259-14.
Musolino, D., de Carli, A., & Massarutto, A. (2017). Evaluation of socio-economic impact of drought events: the case of Po river basin. *European Countryside*, *9*(1), 163-176.
Musolino, D., Massarutto, A., & De Carli, A. (2015). Ex-post evaluation of the socio-economic impacts of drought in some areas in Europe. *Drought: Research and Science-Policy Interfacing, edited by: Andreu, J., Solera, A., Paredes-Arquiola, J., Haro-Monteagudo, D., and van Lanen, HAJ, CRC Press, London*, 71-77.
Nadali, A., Kakhky, E. N. & Nosratabadi, H. E. (2011, April). Evaluating the success level of data mining projects based on CRISP-DM methodology by a Fuzzy expert system. In *Electronics Computer Technology (ICECT), 2011 3rd International Conference on* (Vol. 6, pp. 161-165).IEEE.
Nay, J., Burchfield, E., & Gilligan, J. (2018). A machine-learning approach to forecasting remotely sensed vegetation health. *International journal of remote sensing*, *39*(6), 1800-1816.
Niemeyer, S. (2008). New drought indices. In *Proceedings of the 1st international conference "Drought management: Scientific and technological innovations", Zaragoza, Spain*, pp. 12-14.
Nilsson, N. J. (1996). Introduction to machine learning. An early draft of a proposed textbook.
Oduor, J., Swift, J. and Birch, I. (2014). The Evolution of Kenya's Drought Management System. In *Reducing Disaster: Early Warning Systems for Climate Change*, pp. 209-222. Springer Netherlands.
Ojala, J. (2003, February). Limits and Benefits of Remote Sensing Technology.In *Proceedings California Plant and Soil Conference Farming in Crisis*.

Olang, L. O., Sheffield, J., Wood, E. F., Chaney, N., Guan, K., Sadri, S. and Ogallo, L. (2013). A Drought Monitoring and Forecasting System for Sub-Sahara African Water Resources and Food Security.

Openshaw, S. (1999). Geographical data mining: key design issues. In *Proceedings of GeoComputation* (Vol. 99).

Opitz, D., & Maclin, R. (1999). Popular ensemble methods: An empirical study. *Journal of artificial intelligence research, 11*, 169-198.

Otkin, J. A., Anderson, M. C., Hain, C., Svoboda, M., Johnson, D., Mueller, R., ... & Brown, J. (2016). Assessing the evolution of soil moisture and vegetation conditions during the 2012 United States flash drought. *Agricultural and forest meteorology*, *218*, 230-242.

Oxfam, G.B. (2009) Disaster Risk Reduction in Drought Cycle Management: A Learning Companion. [Online] Available at: https://oxfamilibrary.openrepository.com/bitstream/handle/10546/139094/disaster-risk-reduction-drought-cycle-management-050509-en.pdf?sequence=3&isAllowed=y (Accessed 14 April 2020)

Palmer, W. C. (1965). *Meteorological drought* (Vol. 30). US Department of Commerce, Weather Bureau.

Park, S. (2013). Cloud and cloud shadow effects on the MODIS vegetation index composites of the Korean Peninsula. *International journal of remote sensing*, *34*(4), 1234-1247.

Patruno, J., Fitrzyk, M., & Delgado Blasco, J. M. (2020). Monitoring and Detecting Archaeological Features with Multi-Frequency Polarimetric Analysis. *Remote Sensing*, *12*(1), 1.

Pantuliano, S. & Wekesa M. (2008). *Improving Drought Response in Pastoral Areas of Ethiopia Somali and Afar Regions and Borena Zone of Oromiya Region*. London: ODI.

Peters, A. J., Walter-Shea, E. A., Ji, L., Vina, A., Hayes, M. and Svoboda, M. D. (2002). Drought monitoring with NDVI-based standardized vegetation index.*Photogrammetric engineering and remote sensing*, *68*(1), pp. 71-75.[Online]. Available at: http://asprs.org/a/publications/pers/2002journal/january/2002_jan_71-75.pdf (Accessed 13 November 2015)

Petrakova, A., Affenzeller, M., & Merkurjeva, G. (2015). Heterogeneous versus homogeneous machine learning ensembles. *Information Technology and Management Science*, *18*(1), 135-140.

Phenology (2011). Low or no-cost satellite sensors and data streams utilized for land surface phenology studies: [Online]. Available at: http://phenology.cr.usgs.gov/ndvi_avhrr.php (Accessed 6 November 2014)

Quinlan, J. R. (1996, August). Bagging, boosting, and C4. 5. In *AAAI/IAAI, Vol. 1* (pp. 725-730).

Quiring, S. M. (2009). Developing objective operational definitions for monitoring drought. *Journal of Applied Meteorology and Climatology*, *48*(6), 1217-1229.

Quiring, S. M., & Ganesh, S. (2010). Evaluating the utility of the Vegetation Condition Index (VCI) for monitoring meteorological drought in Texas. *Agricultural and Forest Meteorology*, *150*(3), 330-339.

Ramos, E. G., & Martínez, F. V. (2013). A Review of Artificial Neural Networks: How Well Do They Perform in Forecasting Time Series?. *Analítika: revista de análisis estadístico*, (6), 7-18.

Re, M., & Valentini, G. (2012). 1 Ensemble methods: a review 3.

Rembold, F., Atzberger, C., Savin, I. and Rojas, O. (2013). Using low resolution satellite imagery for yield prediction and yield anomaly detection. *Remote Sensing*, *5*(4), pp. 1704-1733 [Online]. Available at: www.mdpi.com/2072-4292/5/4/1704/pdf (Accessed 19 November 2014)

Rhee, J., & Yang, H. (2018). Drought prediction for areas with sparse monitoring networks: A case study for Fiji. *Water*, *10*(6), 788.

Riedmiller, M. (1994). Advanced supervised learning in multi-layer perceptrons-from backpropagation to adaptive learning algorithms. *Computer standards and interfaces*, *16*(3), 265-278.

Rojas, O., Vrieling, A. & Rembold, F. (2011). Assessing drought probability for agricultural areas in Africa with coarse resolution remote sensing imagery.*RemoteSensing of Environment*, *115*(2), pp. 343-352 [Online]. Available at: http://www.fao.org/climatechange/38004-05a54c6f3e665ea2fabdec165b7948383.pdf (Accessed 2 December 2014)

Running, S., Mu, Q., Zhao, M. (2017). *MOD16A2 MODIS/Terra Net Evapotranspiration 8-Day L4 Global 500m SIN Grid V006* [Data set]. NASA EOSDIS Land Processes DAAC. doi: 10.5067/MODIS/MOD16A2.006

Russom, P. (2011). Big data analytics. *TDWI best practices report, fourth quarter*, *19*(4), 1-34.

Sattari, M. T., Rezazadeh-Joudi, A., & Kusiak, A. (2017). Assessment of different methods for estimation of missing data in precipitation studies. *Hydrology Research*, *48*(4), 1032-1044.

Schaub, S., & Finger, R. (2020, March). Effects of drought on hay and feed grain prices. In *ERL* (Vol. 15, No. 3, p. 034014).

Scheftic, W., Zeng, X., Broxton, P. & Brunke, M. (2014). Intercomparison of seven NDVI products over the United States and Mexico. *Remote Sensing*, *6*(2), 1057-1084 [Online]. Available at: http://www.mdpi.com/2072-4292/6/2/1057 (Accessed 17 November 2014)

Schipper, O. (2003). Declining water resources raise food concerns, *Environmental Science and Technology*, 37(15): 273A-274A.

Scholkopf, B., Smola, A. J., & Bach, F. (2002). *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press.

Shah, H., Rane, V., Nainani, J., Jeyakumar, B., & Giri, N. (2017). Drought Prediction and Management using Big Data Analytics. *Int. J. Comput. Appl*, *162*, 27-30.

Sheffield, J. (2014), *Current and future capabilities for monitoring drought*: [Online]. Available at: http://www.wcrp-climate.org/gdiswksp/presentation_global_drought_monitoring_capabilities_gdapgdis_jpl_dec2014.pdf (Accessed 3 November 2016)

Sheffield, J., Wood, E. F., Lettenmaier, D., & Lipponen, A. (2008). Experimental drought monitoring for Africa. *GEWEX News*, *18*(3), 4-6.

Shirzad, A., Tabesh, M., & Farmani, R. (2014). A comparison between performance of support vector regression and artificial neural network in prediction of pipe burst rate in water distribution networks. *KSCE Journal of Civil Engineering*, *18*(4), 941-948.

Shuaibu, S. M., Ogbodo, J. A., Wasige, E. J., & Mashi, S. A. (2016). Assessing the impact of agricultural drought on maize prices in Kenya with the approach of the SPOT-VEGETATION NDVI remote sensing. *Future of Food: Journal on Food, Agriculture and Society*, *4*(3), 8-18.

Shwartz-Ziv, R., & Tishby, N. (2017). Opening the black box of deep neural networks via information. *arXiv preprint arXiv:1703.00810*.

Silber, J. H., Rosenbaum, P. R., & Ross, R. N. (1995). Comparing the contributions of groups of predictors: which outcomes vary with hospital rather than patient characteristics?. *Journal of the American Statistical Association*, *90*(429), 7-18.

Sivakumar, M. V. K., Wilhite, D. A., Pulwarty, R. S., & Stefanski, R. (2014). The high-level meeting on national drought policy. *Bulletin of the American Meteorological Society*, *95*(4), ES85-ES88.

Sivakumar, M. V., Motha, R. P., Wilhite, D. A., & Qu, J. J. (2011). Towards a compendium on national drought policy. In *Proceedings of the Expert Meeting on the Preparation of a Compendium on National Drought Policy, Washington, DC, USA* (pp. 14-15)

Slonecker, E. T., Shaw, D. M., & Lillesand, T. M. (1998). Emerging legal and ethical issues in advanced remote sensing technology. Photogrammetric engineering and remote sensing, 64(6), 589-595.

Song, Y., Ma, M., & Veroustraete, F. (2010). Comparison and conversion of AVHRR GIMMS and SPOT VEGETATION NDVI data in China. *International Journal of Remote Sensing*, *31*(9), 2377-2392.

Štěpánek, P., Trnka, M., Chuchma, F., Zahradníček, P., Skalák, P., Farda, A., ... & Možný, M. (2018). Drought prediction system for Central Europe and its validation. *Geosciences*, *8*(4), 104.

Su, Z., He, Y., Dong, X., & Wang, L. (2017). Drought monitoring and assessment using remote sensing. In *Remote Sensing of Hydrological Extremes* (pp. 151-172). Springer, Cham.

Sun, L., Mitchell, S. W., & Davidson, A. (2012). Multiple drought indices for agricultural drought risk assessment on the Canadian prairies. *International Journal of Climatology*, *32*(11), 1628-1639.

Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.

Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., ... & Miskus, D. (2002). The drought monitor. *Bulletin of the American Meteorological Society*, *83*(8), 1181-1190.

Swets, D.L., Reed, B.C., Rowland, J.D. & Marko, S.E. (1999). A weighted least-squares approach to temporal NDVI smoothing. In: 1999 ASPRS Annual Conference: From Image to Information, Portland, Oregon, May 17-21.

Proceedings: Bethesda, Maryland, *American Society for Photogrammetry and Remote Sensing*

Tadesse, T., Demisse, G. B., Zaitchik, B., & Dinku, T. (2014). Satellite-based hybrid drought monitoring tool for prediction of vegetation condition in Eastern Africa: A case study for Ethiopia. *Water Resources Research*, *50*(3), 2176-2190.

Tadesse, T., Wardlow, B. D., Hayes, M. J., Svoboda, M. D. and Brown, J. F. (2010). The vegetation outlook (VegOut): A new method for predicting vegetation seasonal greenness. *GIScience& Remote Sensing*, *47*(1), pp. 25-52.

Tarnavsky, E., Grimes, D., Maidment, R., Black, E., Allan, R. P., Stringer, M., ... & Kayitakire, F. (2014). Extension of the TAMSAT satellite-based rainfall monitoring over Africa and from 1983 to present. *Journal of Applied Meteorology and Climatology*, *53*(12), 2805-2822

Thompson, B., Diamond, K. E., McWilliam, R., Snyder, P., & Snyder, S. W. (2005). Evaluating the quality of evidence from correlational research for evidence-based practice. *Exceptional Children*, *71*(2), 181-194.

Tortajada, C., Kastner, M. J., Buurman, J., & Biswas, A. K. (2017). The California drought: Coping responses and resilience building. *Environmental Science & Policy*, *78*, 97-113.

Toté, C., Patricio, D., Boogaard, H., van der Wijngaart, R., Tarnavsky, E., & Funk, C. (2015). Evaluation of satellite rainfall estimates for drought and flood monitoring in Mozambique. *Remote Sensing*, *7*(2), 1758-1776.

Touma, D., Ashfaq, M., Nayak, M. A., Kao, S. C., & Diffenbaugh, N. S. (2015). A multi-model and multi-index evaluation of drought characteristics in the 21st century. *Journal of Hydrology*, *526*, 196-207.

Trambauer, P., Werner, M., Winsemius, H. C., Maskey, S., Dutra, E., & Uhlenbrook, S. (2015). Hydrological drought forecasting and skill assessment for the Limpopo River basin, southern Africa. *Hydrology and Earth System Sciences*, *19*(4), 1695-1711.

Trnka, M., Hayes, M., ek Jurečka, F., Anderson, M., Brázdil, R., Brown, J., ... & Feng, S. (2018). Priority questions in multidisciplinary drought research. *Climate Research*, *75*(3), 241-260.

Tsakiris, G., Nalbantis, I., Vangelis, H., Verbeiren, B., Huysmans, M., Tychon, B., ... & Poelmans, L. (2013). A system-based paradigm of drought analysis for operational management. *Water resources management*, *27*(15), 5281-5297.

Turco, M., Ceglar, A., Prodhomme, C., Soret, A., Toreti, A., & Francisco, J. D. R. (2017). Summer drought predictability over Europe: empirical versus dynamical forecasts. *Environmental Research Letters*, *12*(8), 084006.

Unganai, L. S., and Kogan, F. N. (1998). Drought monitoring and corn yield estimation in Southern Africa from AVHRR data. *Remote Sensing of Environment*, *63*(3), 219-232.

UNISDR (2006). *Global Survey of Early Warning Systems: An assessment of capacities, gaps and opportunities towards building a comprehensive global early warning system for all natural hazards: [*Online]. Available at: http://www.unisdr.org/2006/ppew/info-resources/ewc3/Global-Survey-of-Early-Warning-Systems.pdf (Accessed 8 January 2015)

UNISDR (2009). Drought Risk Reduction Framework and Practices: *Contributing to the Implementation of the Hyogo Framework for Action*. United Nations secretariat of the International Strategy for Disaster Reduction (UNISDR), Geneva, Switzerland, pp. 213.

UNISDR (2012) Number of climate-related disasters, 1980–2011 – Graphic. Available at: http://www.preventionweb.net/english/professional/statistics (Accessed 10 November 2013)

UNISDR, U. (2005). Hyogo framework for action 2005–2015: Building the resilience of nations and communities to disasters. In *Extract from the final report of the World Conference on Disaster Reduction (A/CONF. 206/6)* (Vol. 380). Geneva: The United Nations International Strategy for Disaster Reduction.

UNOOSA (2015), Data application of the month: Drought monitoring. *UN-SPIDER*. [Online]. Available at: http://www.un-spider.org/links-and-resources/data-sources/daotm-drought (Accessed 11 November 2016)

Vicente-Serrano, S. M., Beguería, S., Lorenzo-Lacruz, J., Camarero, J. J., López-Moreno, J. I., Azorin-Molina, C., ... & Sanchez-Lorenzo, A. (2012). Performance of drought indices for ecological, agricultural, and hydrological applications. *Earth Interactions*, *16*(10), 1-27.

Vision 2030, K. (2007): A Globally Competitive and Prosperous Kenya (2007). *Ministry of Planning and National Development and the National Economic and Social Council (NESC), Government of Kenya, Nairobi.*

Wagacha, P. W. (2003). *Support Vector Machines: A critical empirical evaluation* (Unpublished doctoral dissertation). School of Computing & Informatics, University of Nairobi, Nairobi, Kenya

Wan, Z., Hook, S., Hulley, G. (2015). *MOD11A2 MODIS/Terra Land Surface Temperature/Emissivity 8-Day L3 Global 1km SIN Grid V006* [Data set]. NASA EOSDIS LP DAAC. doi: 10.5067/MODIS/MOD11A2.006

Wang, B., Ono, A., Muramatsu, K. and Fujiwara, N. 1999: Automated detection and removal of clouds and their shadows from Landsat TM images. *IEICE Transactions on Information and Systems* E82-D, 453–60.

Wardlow, B. D., Tadesse, T., Brown, J. F., Callahan, K., Swain, S., & Hunt, E. (2012). Vegetation Drought Response Index An Integration of Satellite, Climate, and Biophysical Data.

Wardlow, B.D., & Anderson, M.C. (2012). Remote Sensing of Drought: Innovative Monitoring Approaches (Drought and Water Crises). Broken Sound Parkway: CRC Press.

Wenxia, G., Huanfeng, S., Liangpei, Z. and Wei, G. (2014, March). Normalization of NDVI from Different Sensor System using MODIS Products as Reference. In *IOP Conference Series: Earth and Environmental Science*, 17 (1), p. 1-6.

Wetterhall, F., Winsemius, H. C., Dutra, E., Werner, M., & Pappenberger, E. (2015). Seasonal predictions of agro-meteorological drought indicators for the Limpopo basin. *Hydrology and Earth System Sciences*, *19*(6), 2577-2586.

Wichitarapongsakun, P., Sarin, C., Klomjek, P., & Chuenchooklin, S. (2016). Rainfall prediction and meteorological drought analysis in the Sakae Krang River basin of Thailand. *Agriculture and Natural Resources*, *50*(6), 490-498.

Wilby, R.L., & Dawson, C.W(2013). The statistical downscaling model: Insights from one decade of application, *International Journal of Climatology,* 33(7), pp.1707-1719, ISSN: 0899-8418. (2013)

Wilhite, D. (2006). Drought monitoring and early warning: Concepts, progress and future challenges. *World Meteorological Organization.WMO*, (1006).

Wilhite, D. A. & Glantz, M. IT. (1985) Understanding the drought phenomenon: the role of definition in Environ. Studies 36, pp. 41-45

Wilhite, D. A. (1993). The enigma of drought. In *Drought assessment, management, and planning: Theory and case studies* (pp. 3-15). Springer, Boston, MA.

Wilhite, D. A. (2014). National drought management policy guidelines: A template for action. *Integrated Drought Management Programme (IDMP) Tools and Guidelines Series*, *1*.

Wilhite, D. A., & Svoboda, M. D. (2000). Drought early warning systems in the context of drought preparedness and mitigation. *Early warning systems for drought preparedness and drought management*, 1-21.

Wilhite, D. A., Sivakumar, M. V., & Pulwarty, R. (2014). Managing drought risk in a changing climate: The role of national drought policy. *Weather and Climate Extremes*, *3*, 4-13.

Wilhite, D. A., Sivakumar, M.V.K. &Wood, D.A. (2000) *Drought early warning systems in the context of drought preparedness and mitigation*, World Meteorological Organization (WMO)

Wilhite, D. A., Svoboda, M. D., & Hayes, M. J. (2007). Understanding the complex impacts of drought: A key to enhancing drought mitigation and preparedness. *Water resources management*, *21*(5), 763-774.

Witten, I. H., & Frank, E. (2005). *Data Mining: Practical machine learning tools and techniques*. Burlington: Morgan Kaufmann.

Wolpert, D. H. (2012). What the no free lunch theorems really mean; how to improve search algorithms. In *Santa fe Institute Working Paper* (p. 12).

Wood, E. F., Schubert, S. D., Wood, A. W., Peters-Lidard, C. D., Mo, K. C., Mariotti, A., & Pulwarty, R. S. (2015). Prospects for advancing drought understanding, monitoring, and prediction. *Journal of Hydrometeorology*, *16*(4), 1636-1657.

World Bank (2011). *The drought and food crisis in the Horn of Africa:* Kenya Economic Update.

World Meteorological Organization (WMO). (2012). Standardized precipitation index user guide. [Online]. Available at: http://www.wamis.org/agm/pubs/SPI/WMO_1090_EN.pdf (Accessed 26 April 2019).

Yin, H., Udelhoven, T., Fensholt, R., Pflugmacher, D., & Hostert, P. (2012). How Normalized Difference Vegetation Index (Ndvi) Trends from Advanced Very High Resolution Radiometer (AVHRR) And Système Probatoire d'Observation de la Terre VEGETATION (SPOT VGT) Time Series Differ in Agricultural Areas: An Inner Mongolian Case Study. *Remote Sensing*, *4*(11), 3364-3389.

Yuan, X., Zhang, M., Wang, L., & Zhou, T. (2017). Understanding and seasonal forecasting of hydrological drought in the Anthropocene. *Hydrology and Earth System Sciences*, *21*(11), 5477.

Zambrano, F., Lillo-Saavedra, M., Verbist, K., & Lagos, O. (2016). Sixteen Years of Agricultural Drought Assessment of the BioBío Region in Chile Using a 250 m Resolution Vegetation Condition Index (VCI). *Remote Sensing*, *8*(6), 530.

Zargar, A., Sadiq, R., Naser, B., & Khan, F. I. (2011). A review of drought indices. *Environmental Reviews*, *19*(NA), 333-349.

Zhang, F., Zhang, L. W., Wang, X. Z., & Hung, J. F. (2013). Detecting agro-droughts in Southwest of China using MODIS satellite data. *Journal of Integrative Agriculture*, *12*(1), 159-168.

Zhang, Q., Qi, T., Singh, V. P., Chen, Y. D., & Xiao, M. (2015). Regional frequency analysis of droughts in China: a multivariate perspective. *Water Resources Management*, *29*(6), 1767-1787.

Zhou, Z. H. (2012). *Ensemble methods: foundations and algorithms*. CRC press.

Zhu, Y., Wang, W., Singh, V. P., & Liu, Y. (2016). Combined use of meteorological drought indices at multi-time scales for improving hydrological drought detection. *Science of the Total Environment*, *571*, 1058-1068.

Zikopoulos, P. and Eaton, C. (2011). *Understanding big data: Analytics for enterprise class Hadoop and streaming data*. New York: McGraw-Hill

Appendixes

Appendix A: The getImageWeights function

The *getImageWeights* function is used to convert the VI Usefulness values to corresponging weights for image smoothing.

```
bitShift<-2
bitMask<-15


VIqual   <- raster(ncol=10,nrow=10)
VIqual[] <- bit
viu      <- extractBits(VIqual,bitShift,bitMask)

GetImageWeighs <- function(x,minthres=3,maxthres=7)
{
  maxthres <- maxthres - (minthres-1) # apply shift also to maxthres

  x <- x-minthres # shift vector
  x[x < 0] <- 0
  x[x >= maxthres] <- maxthres

  # create weights
  x <- ((-1) * (x - maxthres))/maxthres
  return(x)
}

#Use of function
weights <- calc(viu,fun= getImageWeights)
plot(weights)
```

Appendix B: Lag-time performance of the models (GAM & ANN)

The study built 102 models using the General Additive Model (GAM) technique. The models were for 1, 2 and 3-month lags on 34 unique model variables. We present the analysis of the lag time performance of the GAM and ANN approaches, assuming the case that the ANN method was also run on the initial set of models as the GAM technique.

1. GAM Model performance by lag time

The comparison of the lag-based performance of the GAM models ordered by their performance in 1-month lag is provided in Figure B1. A summary of the descriptive statistics of the models based on lag time is presented in Figure B1.

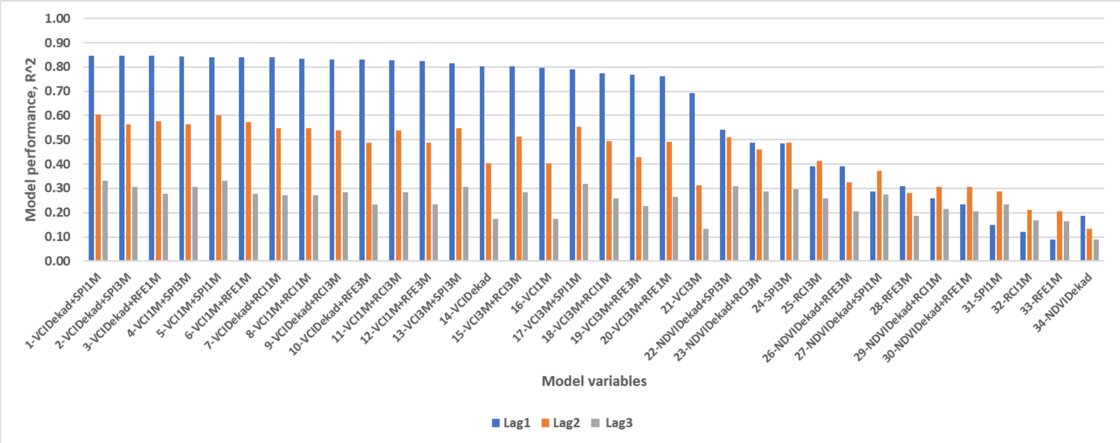


Figure B1. Lag-based performance of the GAM models. The 1-month lags are in blue lines, the 2-month lags in orange lines and 3-month lags in grey.

The models build using 1-month lag variables are shown to perform better than the 2-month and 3-month lags except in 8 out of the 34 cases when 2-month lag time models outperform the 1-month lag models. Even in these 8 cases, the performance of the 2-month lags was still below R^2 of 0.5.

Table B1: Summary of model performance by lag time

Statistic	Lag1	Lag2	Lag3
Mean	0.62	0.44	0.25
Median	0.78	0.49	0.27
Range	0.76	0.47	0.24
Minimum	0.09	0.13	0.09
Maximum	0.85	0.61	0.33

From Table B1, a summary of performance of all the GAM models shows that 1-month prediction has the best performance as compared to the 2-3 months prediction ahead. Despite posting the highest range, 1-month predictions still post a mean performance of R^2=0.62 as compared to 0.44 and 0.25 for 2-month and 3-month lag times, respectively.

2. ANN Model performance by lag time

The study proceeded, in the test of assumptions, to run the ANN process on the entire set of models in the ANN process. A summary of the results is provided following on the same set as the GAM models.

In training, as measured by the performance in the 30% (validation) dataset portion of the training data, the performance of the ANN models is as shown in Figure B2. A summary of the descriptive statistics of the ANN models is provided in Table B2.

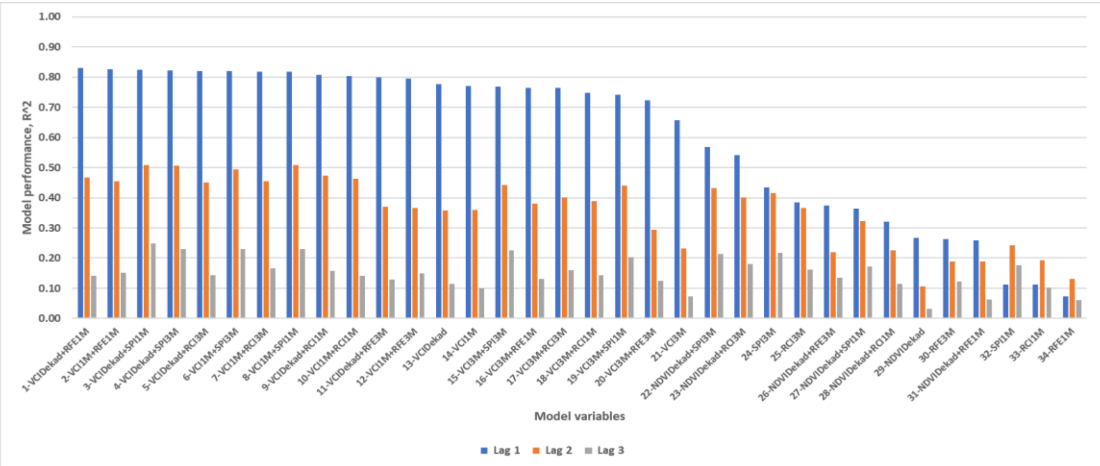


Figure B2. Lag-based performance of the full set (102) ANN models. The 1-month lags are in blue lines, the 2-month lags in orange lines and 3-month lags in grey.

From Figure B2, it is shown that for each of the models, predictions 1-month ahead outperform those for 2-month and 3-month ahead except for the last cases 3 cases (models 32-34) when predictions 2-month ahead are better. At no point does any model have its predictions 3-month ahead out-perform any of the short time period predictions.

Table B2: Summary of model performance by lag time

Statistic	Lag 1	Lag 2	Lag 3
Mean	0.60	0.36	0.15
Median	0.76	0.38	0.15
Range	0.76	0.40	0.22
Minimum	0.07	0.11	0.03
Maximum	0.83	0.51	0.25

The predictions 1-month ahead post the highest range but still end up recording the highest mean of the lagged predictions. At an average R^2 of 0.6 for all the 102 models, the predictions 1-month ahead are judged predictive enough for use in an operational ex-ante system. The best model for prediction 1-month ahead differs from the best model for 2-month and 3-month ahead predictions. Both models have the variable VCIDekad while RFE for the predictions 1-month ahead and SPI1M for both predictions 2-months and 3-months ahead.

Appendix C: Full-list of the performance of the pre-study GAM and ANN models

The full list of GAM models is presented in Table C3 while that of ANN models is presented in Table C4 respectively.

Table C3. GAM models in decreasing order of R^2 with the overfit index provided.

No	Model	R^2 Training	R^2 Validation	Overfit Index	Overfit	Lag Time
1	VCIDekad_lag1+SPI1M_lag1	0.86	0.85	0.01	No	Lag1
2	VCIDekad_lag1+SPI3M_lag1	0.86	0.85	0.01	No	Lag1
3	VCIDekad_lag1+RFE1M_lag1	0.85	0.85	0.01	No	Lag1
4	VCI1M_lag1+SPI3M_lag1	0.85	0.84	0.01	No	Lag1
5	VCI1M_lag1+SPI1M_lag1	0.85	0.84	0.01	No	Lag1
6	VCI1M_lag1+RFE1M_lag1	0.85	0.84	0.01	No	Lag1
7	VCIDekad_lag1+RCI1M_lag1	0.85	0.84	0.01	No	Lag1
8	VCI1M_lag1+RCI1M_lag1	0.84	0.83	0.01	No	Lag1

9	VCIDekad_lag1+RCI3M_lag1	0.84	0.83	0.01	No	Lag1
10	VCIDekad_lag1+RFE3M_lag1	0.84	0.83	0.01	No	Lag1
11	VCI1M_lag1+RCI3M_lag1	0.84	0.83	0.01	No	Lag1
12	VCI1M_lag1+RFE3M_lag1	0.83	0.83	0.01	No	Lag1
13	VCI3M_lag1+SPI3M_lag1	0.82	0.82	0.01	No	Lag1
14	VCIDekad_lag1	0.81	0.8	0.01	No	Lag1
15	VCI3M_lag1+RCI3M_lag1	0.81	0.8	0.01	No	Lag1
16	VCI1M_lag1	0.81	0.8	0.01	No	Lag1
17	VCI3M_lag1+SPI1M_lag1	0.81	0.79	0.01	No	Lag1
18	VCI3M_lag1+RCI1M_lag1	0.78	0.77	0.01	No	Lag1
19	VCI3M_lag1+RFE3M_lag1	0.78	0.77	0.01	No	Lag1
20	VCI3M_lag1+RFE1M_lag1	0.78	0.76	0.01	No	Lag1
21	VCI3M_lag1	0.72	0.69	0.02	No	Lag1
22	VCIDekad_lag2+SPI1M_lag2	0.61	0.61	0	No	Lag2
23	VCI1M_lag2+SPI1M_lag2	0.6	0.6	0	No	Lag2
24	VCIDekad_lag2+RFE1M_lag2	0.58	0.58	0	No	Lag2
25	VCI1M_lag2+RFE1M_lag2	0.58	0.57	0	No	Lag2
26	VCI1M_lag2+SPI3M_lag2	0.57	0.56	0.01	No	Lag2
27	VCIDekad_lag2+SPI3M_lag2	0.57	0.56	0.01	No	Lag2
28	VCI3M_lag2+SPI1M_lag2	0.56	0.56	0	No	Lag2
29	VCIDekad_lag2+RCI1M_lag2	0.56	0.55	0.02	No	Lag2
30	VCI3M_lag2+SPI3M_lag2	0.55	0.55	0	No	Lag2
31	VCI1M_lag2+RCI1M_lag2	0.56	0.55	0.02	No	Lag2
32	NDVIDekad_lag1+SPI3M_lag1	0.56	0.54	0.02	No	Lag1
33	VCIDekad_lag2+RCI3M_lag2	0.55	0.54	0.02	No	Lag2
34	VCI1M_lag2+RCI3M_lag2	0.55	0.54	0.02	No	Lag2
35	VCI3M_lag2+RCI3M_lag2	0.53	0.51	0.01	No	Lag2
36	NDVIDekad_lag2+SPI3M_lag2	0.52	0.51	0.01	No	Lag2
37	VCI3M_lag2+RCI1M_lag2	0.51	0.49	0.02	No	Lag2
38	VCI3M_lag2+RFE1M_lag2	0.5	0.49	0.01	No	Lag2
39	NDVIDekad_lag1+RCI3M_lag1	0.51	0.49	0.02	No	Lag1
40	VCI1M_lag2+RFE3M_lag2	0.51	0.49	0.02	No	Lag2
41	VCIDekad_lag2+RFE3M_lag2	0.51	0.49	0.02	No	Lag2
42	SPI3M_lag2	0.49	0.49	0	No	Lag2
43	SPI3M_lag1	0.5	0.48	0.02	No	Lag1
44	NDVIDekad_lag2+RCI3M_lag2	0.48	0.46	0.02	No	Lag2
45	VCI3M_lag2+RFE3M_lag2	0.44	0.43	0.02	No	Lag2
46	RCI3M_lag2	0.42	0.41	0.01	No	Lag2
47	VCI1M_lag2	0.43	0.4	0.03	No	Lag2
48	VCIDekad_lag2	0.43	0.4	0.03	No	Lag2
49	NDVIDekad_lag1+RFE3M_lag1	0.41	0.39	0.02	No	Lag1
50	RCI3M_lag1	0.41	0.39	0.02	No	Lag1
51	NDVIDekad_lag2+SPI1M_lag2	0.4	0.37	0.03	No	Lag2

52	VCIDekad_lag3+SPI1M_lag3	0.35	0.33	0.01	No	Lag3
53	VCI1M_lag3+SPI1M_lag3	0.34	0.33	0.01	No	Lag3
54	NDVIDekad_lag2+RFE3M_lag2	0.35	0.33	0.02	No	Lag2
55	VCI3M_lag3+SPI1M_lag3	0.33	0.32	0.01	No	Lag3
56	VCI3M_lag2	0.33	0.31	0.02	No	Lag2
57	RFE3M_lag1	0.32	0.31	0.01	No	Lag1
58	NDVIDekad_lag3+SPI3M_lag3	0.33	0.31	0.02	No	Lag3
59*	NDVIDekad_lag2+RCI1M_lag2	0.35	0.31	0.05	Yes	Lag2
60	VCI1M_lag3+SPI3M_lag3	0.32	0.31	0.02	No	Lag3
61	VCI3M_lag3+SPI3M_lag3	0.32	0.31	0.01	No	Lag3
62	VCIDekad_lag3+SPI3M_lag3	0.32	0.31	0.01	No	Lag3
63	NDVIDekad_lag2+RFE1M_lag2	0.34	0.31	0.03	No	Lag2
64	SPI3M_lag3	0.31	0.3	0.02	No	Lag3
65	SPI1M_lag2	0.32	0.29	0.03	No	Lag2
66	NDVIDekad_lag3+RCI3M_lag3	0.31	0.29	0.02	No	Lag3
67	NDVIDekad_lag1+SPI1M_lag1	0.31	0.29	0.03	No	Lag1
68	VCI1M_lag3+RCI3M_lag3	0.3	0.28	0.02	No	Lag3
69	VCI3M_lag3+RCI3M_lag3	0.3	0.28	0.02	No	Lag3
70	VCIDekad_lag3+RCI3M_lag3	0.3	0.28	0.02	No	Lag3
71	RFE3M_lag2	0.29	0.28	0.01	No	Lag2
72	VCIDekad_lag3+RFE1M_lag3	0.31	0.28	0.03	No	Lag3
73	VCI1M_lag3+RFE1M_lag3	0.31	0.28	0.03	No	Lag3
74	NDVIDekad_lag3+SPI1M_lag3	0.3	0.28	0.02	No	Lag3
75	VCIDekad_lag3+RCI1M_lag3	0.29	0.27	0.02	No	Lag3
76	VCI1M_lag3+RCI1M_lag3	0.29	0.27	0.02	No	Lag3
77	VCI3M_lag3+RFE1M_lag3	0.3	0.27	0.03	No	Lag3
78	VCI3M_lag3+RCI1M_lag3	0.28	0.26	0.02	No	Lag3
79	RCI3M_lag3	0.28	0.26	0.02	No	Lag3
80	NDVIDekad_lag1+RCI1M_lag1	0.28	0.26	0.02	No	Lag1
81	VCIDekad_lag3+RFE3M_lag3	0.25	0.24	0.01	No	Lag3
82	VCI1M_lag3+RFE3M_lag3	0.25	0.23	0.01	No	Lag3
83	NDVIDekad_lag1+RFE1M_lag1	0.26	0.23	0.02	No	Lag1
84	SPI1M_lag3	0.25	0.23	0.02	No	Lag3
85	VCI3M_lag3+RFE3M_lag3	0.24	0.23	0.02	No	Lag3
86	NDVIDekad_lag3+RCI1M_lag3	0.24	0.22	0.02	No	Lag3
87*	RCI1M_lag2	0.25	0.21	0.04	Yes	Lag2
88	RFE1M_lag2	0.24	0.21	0.03	No	Lag2
89	NDVIDekad_lag3+RFE1M_lag3	0.24	0.21	0.03	No	Lag3
90	NDVIDekad_lag3+RFE3M_lag3	0.23	0.2	0.02	No	Lag3
91	RFE3M_lag3	0.21	0.19	0.02	No	Lag3
92	NDVIDekad_lag1	0.22	0.19	0.03	No	Lag1
93	VCI1M_lag3	0.19	0.18	0.01	No	Lag3
94	VCIDekad_lag3	0.19	0.18	0.01	No	Lag3

95	RCI1M_lag3	0.19	0.17	0.02	No	Lag3
96	RFE1M_lag3	0.2	0.17	0.03	No	Lag3
97	SPI1M_lag1	0.17	0.15	0.03	No	Lag1
98	VCI3M_lag3	0.15	0.13	0.02	No	Lag3
99	NDVIDekad_lag2	0.16	0.13	0.03	No	Lag2
100	RCI1M_lag1	0.13	0.12	0.01	No	Lag1
101	RFE1M_lag1	0.11	0.09	0.02	No	Lag1
102	NDVIDekad_lag3	0.11	0.09	0.02	No	Lag3

[1] The overfit models are marked with * on the column No.

The GAM models have only 2 out of 102 (under 2%) of models judged as overfitting. This is as compared to the ANN models in Table C4 that indicates 64 out of 102 models losing performance in validation by more than an R^2 of 0.03 as compared to their performance in training.

Table C4. ANN models in decreasing order of R^2 with the overfit index provided.

No	Model	R^2 Training	R^2 Validation	Overfit Index	Overfit	Lag Time
1	VCIDekad_lag1+RFE1M_lag1	0.84	0.83	0.01	No	1
2	VCI1M_lag1+RFE1M_lag1	0.84	0.83	0.01	No	1
3	VCIDekad_lag1+SPI1M_lag1	0.84	0.82	0.02	No	1
4	VCIDekad_lag1+SPI3M_lag1	0.84	0.82	0.02	No	1
5	VCIDekad_lag1+RCI3M_lag1	0.84	0.82	0.02	No	1
6	VCI1M_lag1+SPI3M_lag1	0.84	0.82	0.02	No	1
7	VCI1M_lag1+RCI3M_lag1	0.84	0.82	0.02	No	1
8	VCI1M_lag1+SPI1M_lag1	0.84	0.82	0.02	No	1
9	VCIDekad_lag1+RCI1M_lag1	0.82	0.81	0.02	No	1
10	VCI1M_lag1+RCI1M_lag1	0.82	0.80	0.02	No	1
11	VCIDekad_lag1+RFE3M_lag1	0.82	0.80	0.02	No	1
12	VCI1M_lag1+RFE3M_lag1	0.81	0.79	0.02	No	1
13	VCIDekad_lag1	0.79	0.78	0.01	No	1
14	VCI1M_lag1	0.78	0.77	0.01	No	1
15	VCI3M_lag1+SPI3M_lag1	0.79	0.77	0.03	No	1
16	VCI3M_lag1+RFE1M_lag1	0.77	0.77	0.01	No	1
17	VCI3M_lag1+RCI3M_lag1	0.79	0.76	0.03	No	1
18	VCI3M_lag1+RCI1M_lag1	0.77	0.75	0.02	No	1
19*	VCI3M_lag1+SPI1M_lag1	0.78	0.74	0.04	Yes	1
20	VCI3M_lag1+RFE3M_lag1	0.74	0.72	0.02	No	1
21	VCI3M_lag1	0.68	0.66	0.02	No	1
22*	NDVIDekad_lag1+SPI3M_lag1	0.60	0.57	0.04	Yes	1
23*	NDVIDekad_lag1+RCI3M_lag1	0.59	0.54	0.05	Yes	1
24*	VCI1M_lag2+SPI1M_lag2	0.57	0.51	0.06	Yes	2
25*	VCIDekad_lag2+SPI1M_lag2	0.58	0.51	0.07	Yes	2
26*	VCIDekad_lag2+SPI3M_lag2	0.54	0.51	0.04	Yes	2
27*	VCI1M_lag2+SPI3M_lag2	0.56	0.49	0.07	Yes	2

28*	VCIDekad_lag2+RCI1M_lag2	0.53	0.47	0.06	Yes	2
29*	VCIDekad_lag2+RFE1M_lag2	0.52	0.47	0.06	Yes	2
30*	VCI1M_lag2+RCI1M_lag2	0.53	0.46	0.07	Yes	2
31*	VCI1M_lag2+RCI3M_lag2	0.53	0.46	0.08	Yes	2
32*	VCI1M_lag2+RFE1M_lag2	0.53	0.46	0.07	Yes	2
33*	VCIDekad_lag2+RCI3M_lag2	0.52	0.45	0.07	Yes	2
34*	VCI3M_lag2+SPI3M_lag2	0.52	0.44	0.08	Yes	2
35*	VCI3M_lag2+SPI1M_lag2	0.50	0.44	0.06	Yes	2
36*	SPI3M_lag1	0.47	0.43	0.03	Yes	1
37*	NDVIDekad_lag2+SPI3M_lag2	0.48	0.43	0.05	Yes	2
38	SPI3M_lag2	0.42	0.42	0.00	No	2
39*	VCI3M_lag2+RCI3M_lag2	0.49	0.40	0.09	Yes	2
40*	NDVIDekad_lag2+RCI3M_lag2	0.45	0.40	0.05	Yes	2
41*	VCI3M_lag2+RCI1M_lag2	0.51	0.39	0.12	Yes	2
42*	RCI3M_lag1	0.43	0.39	0.04	Yes	1
43*	VCI3M_lag2+RFE1M_lag2	0.47	0.38	0.09	Yes	2
44*	NDVIDekad_lag1+RFE3M_lag1	0.47	0.37	0.09	Yes	1
45*	VCIDekad_lag2+RFE3M_lag2	0.46	0.37	0.09	Yes	2
46*	VCI1M_lag2+RFE3M_lag2	0.46	0.37	0.09	Yes	2
47	RCI3M_lag2	0.38	0.37	0.01	No	2
48*	NDVIDekad_lag1+SPI1M_lag1	0.43	0.36	0.06	Yes	1
49*	VCI1M_lag2	0.39	0.36	0.03	Yes	2
50*	VCIDekad_lag2	0.39	0.36	0.03	Yes	2
51*	NDVIDekad_lag2+SPI1M_lag2	0.39	0.32	0.07	Yes	2
52*	NDVIDekad_lag1+RCI1M_lag1	0.35	0.32	0.03	Yes	1
53*	VCI3M_lag2+RFE3M_lag2	0.41	0.29	0.12	Yes	2
54	NDVIDekad_lag1	0.28	0.27	0.01	No	1
55	RFE3M_lag1	0.27	0.26	0.01	No	1
56*	NDVIDekad_lag1+RFE1M_lag1	0.34	0.26	0.08	Yes	1
57*	VCIDekad_lag3+SPI1M_lag3	0.31	0.25	0.06	Yes	3
58	SPI1M_lag2	0.26	0.24	0.02	No	2
59*	VCI3M_lag2	0.28	0.23	0.05	Yes	2
60*	VCIDekad_lag3+SPI3M_lag3	0.30	0.23	0.07	Yes	3
61*	VCI1M_lag3+SPI1M_lag3	0.31	0.23	0.08	Yes	3
62*	VCI1M_lag3+SPI3M_lag3	0.31	0.23	0.08	Yes	3
63*	NDVIDekad_lag2+RCI1M_lag2	0.31	0.23	0.09	Yes	2
64*	VCI3M_lag3+SPI3M_lag3	0.28	0.23	0.06	Yes	3
65*	NDVIDekad_lag2+RFE3M_lag2	0.31	0.22	0.10	Yes	2
66	SPI3M_lag3	0.23	0.22	0.01	No	3
67*	NDVIDekad_lag3+SPI3M_lag3	0.27	0.21	0.06	Yes	3
68*	VCI3M_lag3+SPI1M_lag3	0.32	0.20	0.12	Yes	3
69	RCI1M_lag2	0.20	0.19	0.01	No	2
70*	NDVIDekad_lag2+RFE1M_lag2	0.24	0.19	0.05	Yes	2
71*	RFE3M_lag2	0.23	0.19	0.05	Yes	2
72*	NDVIDekad_lag3+RCI3M_lag3	0.27	0.18	0.09	Yes	3
73	SPI1M_lag3	0.20	0.18	0.02	No	3
74*	NDVIDekad_lag3+SPI1M_lag3	0.27	0.17	0.10	Yes	3
75*	VCI1M_lag3+RCI3M_lag3	0.25	0.17	0.08	Yes	3

76*	RCI3M_lag3	0.20	0.16	0.04	Yes	3
77*	VCI3M_lag3+RCI3M_lag3	0.27	0.16	0.11	Yes	3
78*	VCIDekad_lag3+RCI1M_lag3	0.27	0.16	0.11	Yes	3
79*	VCI1M_lag3+RFE1M_lag3	0.23	0.15	0.07	Yes	3
80*	VCI1M_lag3+RFE3M_lag3	0.21	0.15	0.06	Yes	3
81*	VCI3M_lag3+RCI1M_lag3	0.30	0.14	0.15	Yes	3
82*	VCIDekad_lag3+RCI3M_lag3	0.27	0.14	0.12	Yes	3
83*	VCI1M_lag3+RCI1M_lag3	0.30	0.14	0.16	Yes	3
84*	VCIDekad_lag3+RFE1M_lag3	0.24	0.14	0.10	Yes	3
85*	NDVIDekad_lag3+RFE3M_lag3	0.19	0.13	0.06	Yes	3
86	RFE1M_lag2	0.14	0.13	0.01	No	2
87*	VCI3M_lag3+RFE1M_lag3	0.20	0.13	0.07	Yes	3
88*	VCIDekad_lag3+RFE3M_lag3	0.22	0.13	0.09	Yes	3
89*	VCI3M_lag3+RFE3M_lag3	0.19	0.12	0.07	Yes	3
90	RFE3M_lag3	0.14	0.12	0.01	No	3
91	VCIDekad_lag3	0.14	0.11	0.03	No	3
92*	NDVIDekad_lag3+RCI1M_lag3	0.18	0.11	0.07	Yes	3
93	SPI1M_lag1	0.14	0.11	0.02	No	1
94	RCI1M_lag1	0.11	0.11	(0.00)	No	1
95	NDVIDekad_lag2	0.13	0.11	0.02	No	2
96*	RCI1M_lag3	0.13	0.10	0.03	Yes	3
97*	VCI1M_lag3	0.15	0.10	0.05	Yes	3
98	VCI3M_lag3	0.09	0.07	0.02	No	3
99	RFE1M_lag1	0.07	0.07	(0.00)	No	1
100*	NDVIDekad_lag3+RFE1M_lag3	0.14	0.06	0.08	Yes	3
101	RFE1M_lag3	0.08	0.06	0.02	No	3
102	NDVIDekad_lag3	0.05	0.03	0.01	No	3